DOCUMENT RESUME

ED 417 346 CE 076 101

AUTHOR Green, Muriel; Bartram, Dave
TITLE Initial Assessment to Identify Learning Needs.
INSTITUTION Further Education Development Agency, London (England).
ISSN ISSN-1361-9977
PUB DATE 1998-00-00
NOTE 51p.
AVAILABLE FROM Further Education Development Agency, Publications Dept., Coombe Lodge, Blagdon, Bristol BS40 7RG, United Kingdom (7.50 pounds).
PUB TYPE Collected Works - Serials (022) -- Guides - Non-Classroom (055)
JOURNAL CIT FE Matters; v2 n7 1998
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Adult Programs; *Adult Students; Diagnostic Tests; *Educational Needs; *Educational Practices; *Evaluation Methods; Foreign Countries; National Standards; *Needs Assessment; Objective Tests; Postsecondary Education; *Student Evaluation; Technical Institutes; Test Construction; Vocational Education
IDENTIFIERS *United Kingdom

ABSTRACT

This document, which is intended for curriculum managers, student services managers, and learning support managers in further education (FE) colleges throughout the United Kingdom, presents the practical experiences of FE colleges that have implemented a range of approaches to initial assessment and offers technical information, good practice criteria, guidance, and advice regarding developing and using initial assessment materials. The following are among the topics discussed: state of initial assessment in FE colleges; approaches to initial assessment (understanding the rationale for initial assessment, managing initial assessment in FE colleges, choosing and/or devising assessment materials, timing initial assessment, grading and feedback); technical aspects of initial assessment (assessment and testing, assessment policy, definition of initial assessment, assessment for placement versus classification, selection interviews, screening and diagnosis, rationale for diagnostic testing, and use of screening tests and diagnostic assessments); tools and techniques (quality criteria, required resources, key points); advice to colleges (developing an assessment policy; making contracts; understanding options; and deciding whether to test, which tests to use; and whether to use internally developed tests or tests from other colleges). Appended are the following: draft code of good practice testing; guidelines for defining objective tests; and information about designing and developing diagnostic tests. (MN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 417 346

FEDA Further Education Development Agency

Initial assessment to identify learning needs

Muriel Green and Dave Bartram

FE matters

CE036101

BEST COPY AVAILABLE

2

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

S. Browne

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

FEDA Further Education Development Agency

Initial assessment to identify learning needs

Muriel Green and Dave Bartram

FE matters

3

Published by the **Further Education Development Agency** (FEDA), Dumbarton House, 68 Oxford Street, London W1N 0DA
Tel: [0171] 436 0020 Fax: [0171] 436 0349

Feedback and orders should be directed to:
Publications Department, FEDA,
Coombe Lodge, Blagdon, Bristol BS40 7RG
Tel: [01761] 462 503 Fax: [01761] 463 140

Registered with the Charity Commissioners

Editor: **Jennifer Rhys**

Designer: **Mike Pope**

Printed by: **Blackmore Limited, Shaftesbury**

Cover photograph: **Reproduced with kind permission from GLOSCAT**

ISSN: 1361-9977

© 1998 FEDA

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, electrical, chemical, optical, photocopying, recording or otherwise, without prior permission of the copyright owner, except as follows:

FEDA grants the purchaser a non-transferable licence to use any case studies and evaluation frameworks as follows: (i) they may be photocopied and used as many times as required, within a single site in the purchasing institution solely; (ii) short excerpts from them may be incorporated into any development paper or review document if the source is acknowledged and the document is not offered for sale; (iii) permission for other uses should be sought from FEDA, Blagdon.

Acknowledgements

FEDA thanks the following colleges that contributed to our work on initial assessment practice.

- Arnold and Carlton College
- Basford Hall College
- Basingstoke College of Technology
- Bilborough Sixth Form College
- Calderdale College
- Carmarthenshire College
- Clarendon College
- Gloucestershire College of Arts and Technology
- Hartlepool College
- High Pavement Sixth Form College
- Hull College
- Mid Kent College of Higher and Further Education
- Milton Keynes College
- North Oxfordshire College and School of Art
- Rotherham College of Arts and Technology
- Sandwell College of Further and Higher Education
- South Nottingham College
- Southwark College
- The People's College
- The Sheffield College
- West Suffolk College

About the authors

Muriel Green has had a range of roles in all sectors of education. She joined the former FEU from an adviser/inspector role in Leicestershire where she had been a member of the LEA's assessment team. Her work has focused particularly on learner and learning support and she is currently managing a QCA-funded project on initial assessment, again working with Dave Bartram. Her recent publications include *Additional support, retention and guidance in urban colleges* (1997) and *Different approaches to learning support* (1998).

Dave Bartram is Professor of Psychology and Dean of the Faculty of Science and the Environment at the University of Hull. He is a Chartered Occupational Psychologist, Fellow of the British Psychological Society, and a Fellow of the Ergonomics Society. He is the author of over 200 journal articles, technical reports, books and book chapters, most of which concern assessment and assesssment issues, particularly in the context of personnel selection. He has also been responsible for the design and implementation of a range of assessment software products and paper-and-pencil tests which are used in a number of countries.

Contents

Foreword 4

Summary 5

1 Introduction 7
Background
Initial assessment in FE colleges

2 Approaches to initial assessment 9
Why offer initial assessment?
Managing initial assessment in FE colleges
Choosing and/or devising assessment materials
Timing initial assessment
Administering initial assessment
Marking and feedback

3 Technical aspects of initial assessment 25
Assessment and testing
Assessment policy
Initial assessment: clarification and definitions
Selection: placement or classification
The selection interview
Screening and diagnosis
Why do we need diagnostic testing?
Using screening tests and diagnostic assessments

4 Tools and techniques 32
Quality criteria
What resources are required?
Key points

5 Advice to colleges: dos and don'ts 38
Assessment policy
Making a contract
What are the options?
To test or not to test
Which tests to use
Do-it-yourself?

6 Conclusions 39

Appendices
1 Draft code of good practice testing 40
2 Defining objective tests 41
3 How are diagnostic tests designed and developed? 43

References 46

Foreword

This book focuses on initial assessment practice in further education colleges and has been written for curriculum managers, student services managers and learning support managers. It may also be of use to schools, careers services, training providers and higher education.

Using the outcomes of two related FEDA initial assessment projects, it presents colleges' practical approaches to initial assessment and offers technical information, good practice criteria, guidance and advice based on an evaluation of college materials.

I am grateful to all members of the earlier project's team, Sally Faraday, Pho Kypri and Anna Reisenberger, who worked with me to support project colleges in this important area of work. The commitment, energy and enthusiasm of the colleges themselves must be acknowledged as must the professionalism of those college representatives who have already done so much to disseminate their experience via FEDA conferences.

I am particularly grateful to the colleges which were brave enough, in our most recent project, to submit examples of their 'home-grown' materials, to be scrutinised and evaluated by our expert consultant, David Bartram. Dave Bartram is a Chartered Occupational Psychologist and Professor of Psychology at Hull University, with an international reputation for his work on assessment. I have very much appreciated his expertise and support in recent and current development work. I hope that readers find his contribution to this report, as co-author, as informative and useful as we have.

Muriel Green
FEDA education staff

Summary

Background

This book focuses on initial assessment practice in further education (FE) colleges. It has been written for curriculum managers, student services managers and learning support managers. It may also be of use to schools, careers services, training providers and higher education.

Using the outcomes of two related FEDA initial assessment projects (Initial Assessment and Learning Support) it presents the practical experiences of colleges which implemented a range of approaches to initial assessment and offers technical information, good practice criteria, guidance and advice based on an evaluation of the initial assessment materials developed in selected colleges.

Best practice

Best practice in the management and implementation of initial assessment includes:

- a common and shared understanding of the purpose of initial assessment
- a clear management strategy which reflects the college mission and purpose
- an assessment policy and code of practice
- a transparent management structure which identifies roles and responsibilities, clear lines of communication and accountability.

In reviewing the management and implementation of initial assessment, colleges will need to reflect on the purposes of initial assessment and on the extent to which existing policies, structures, roles and responsibilities will help the college achieve in relation to each of the identified purposes of initial assessment.

Purpose and timing

The purpose and timing of assessment will need to be considered so that:

- assessment which seeks to aid placement takes place pre-entry
- assessment which seeks to identify skill levels and support takes place at entry
- assessment which seeks to identify the specific nature of support needs is an integral part of the induction process for those students who have been identified by screening as needing support.

Administering initial assessment

Colleges need to:

- develop clear implementation strategies which relate to the purpose of initial assessment, in line with the college's assessment policy
- clarify and confirm the roles and responsibilities of all staff involved in the process
- establish clear and simple systems for managing the process of initial assessment
- offer training, guidance and support to staff to ensure they feel comfortable and confident in their roles
- communicate clearly to students the purposes and processes of initial assessment.

College materials

FEDA has collected examples of home-grown diagnostic assessment tools developed by cutting edge colleges, including a few which sell their assessment materials to other colleges. Problems with college-developed materials included:

- setting cut-off scores for classifying people into different levels of attainment without any empirical justification for the location of the cut-off points
- inappropriate diagnostic interpretation of responses to individual items or small subsets of items on tests designed as screening tests
- lack of evidence of reliability or validity or other supporting technical documentation
- application of inappropriate 'summative' educational assessment approaches to diagnostic assessment.

The publication

This publication presents good practice criteria and guidance to support colleges in producing more rigorous and robust assessment instruments.

In choosing or devising initial assessment materials colleges need to:

- be clear about the purpose for which tests will be used
- identify the skill demands of particular programmes and ensure that initial assessment materials relate to them
- ensure that specialist learning support staff can be available to advise, guide or work with curriculum teams
- clarify roles and responsibilities in line with college policy and a code of practice.

Quality control in the use of objective assessment depends on the combination of robust, relevant instruments and competent users. Six main areas of quality control criteria are highlighted: **scope**, **reliability**, **validity**, **fairness**, **acceptability** and **practicality**. To judge the overall cost-effectiveness of using any assessment method, one should evaluate it against all six of these criteria.

Key messages

- Accurate diagnosis is not the same as effective 'treatment' – but it does provide the information needed to target support resources more efficiently.
- Overall efficiency requires accurate diagnosis combined with appropriate placement and good learning support.
- Poor diagnosis is costly to develop and can lead to a mis-direction and waste of support effort.
- Poor diagnosis can end up costing you more than no diagnosis.

1 Introduction

Background

FEFC's report, *Measuring achievement* (1997) provides interesting data on student continuation and achievement rates. The continuation rate for the median college in 1994-5 was 88% with achievement at 71%. The average college achievement rates varied by college type, from 82% in specialist colleges to 63% in general FE and tertiary. Range of achievement was much wider than continuation rates – a quarter of FE colleges had achievement rates of less than 51%.

Colleges are thus keen to address problems of retention and achievement and are giving a high priority to effective learning and teaching. They are attaching great importance to the need to make early and informed judgements about learners' experience and skills in order to guide them towards and select them for the most appropriate programme. They are also keen to ensure that learners are able to manage their work, make personal and academic progress and achieve their learning goals.

The report from the FEFC's Learning Difficulties and/or Disabilities Committee led by Professor Tomlinson, *Inclusive Learning* (1996), promotes an approach to learning 'which we would want to see everywhere':

> *At the heart of our thinking lies the idea of match or fit between how the learner learns best, what they need and want to learn, and what is required from the sector, a college and teachers for successful learning to take place.*
>
> *(Tomlinson, 1996 2.3)*

Early and effective assessment of students' requirements is critical to the concept of inclusive learning. Chapter 5 of *Inclusive Learning* sets out how students' requirements will need to be assessed in order to ensure inclusive learning. While acknowledging the desirability of inclusive learning and the particular assessment requirements of individuals with learning difficulties and disabilities, this book focuses on the whole-college approaches to initial assessment which were going on at the same time as the FEFC's Committee was examining educational provision for those with learning difficulties and/or disabilities. In sharing information on examples of initial assessment practice, from which we have distilled general guidance and good practice criteria, FEDA seeks to contribute to improved initial assessment processes and to the development of more 'inclusive' colleges.

It is important to signal at this early stage that initial assessment on its own will do little to ensure effective learning. It is the way in which learners, teachers and institutions are able to use the information generated that is critical to the learners' success. FEDA's publication, *Different approaches to learning support* (Green, 1998) disseminates good practice and offers advice on effective support systems. Readers seeking information, advice and guidance to help take forward their own provision will find it useful to work with both publications together.

Initial assessment in FE colleges

There seems to be some confusion about initial assessment practices. Initial assessment is the first experience of a sequence of assessment processes which enables an assessor to make judgements about the assessed. Sometimes learners will make judgements about themselves, measuring their performance against shared criteria through a self-assessment process. Assessment in an educational context should support learners and learning. For learners it will:

- identify what has been learned
- provide feedback
- identify what still needs to to be learned
- enable learners to set targets which ensure success in new learning
- allow learners to take responsibility for personal development.

For lecturers, assessment will:

- provide confirmation of what has been learned and what still needs to be learned
- be a basis for discussion with students and other staff
- help with evaluation and planning of programme design and delivery.

Assessment can be both forward-looking and backward-looking. Backward-looking assessment measures what has been learned or achieved. Forward-looking assessment looks at what may be achieved or learned in the future. Most educational assessment is backward-looking: it attempts to measure what has been achieved and to provide some form of credit for it (in the form of whole or part qualifications).

Initial assessment has both forward- and backward-looking components. It is concerned both with assessing where the student is now and with making judgements about his or her capacity to progress along one or other of a number of paths. It is the latter aspect of assessment which creates the greatest difficulty and demands careful consideration of the technical nature of the process.

A later section of this book goes into the technicalities of assessment in some depth. However, we need to signal briefly at this early stage that initial assessment in colleges is a complex set of processes which can:

- inform guidance
- facilitate selection and placement
- screen for levels of literacy and numeracy
- diagnose programme specific learning needs.

Screening and diagnostic assessment are different but related processes. Page 27 of this publication seeks to clarify the distinction between these two forms of initial assessment.

Colleges using initial assessment processes including screening and diagnosis, do so for different purposes and use different approaches. The next chapter presents an overview of college practice and makes recommendations which are expanded in later sections of the documents.

2 Approaches to initial assessment

Why offer initial assessment?

National imperatives to recruit and retain more students are seen as the major driving forces behind initial assessment developments. Colleges keen to recruit more, and different, learners recognise that growth inevitably leads to a change in the student profile, a possible increase in the numbers of non-traditional students and a need to work harder to support students so that they are able to make positive progress toward their stated learning goals.

This section includes information gained from FEDA action research with a small group of colleges (17), selected to represent the diversity of FE provision across the country looking in detail at the different approaches to initial assessment.

While colleges recognise the value of initial assessment for its own sake, the Funding Methodology, with additional support units, is flagged up by project colleges as the most powerful lever for change in initial assessment practice.

The need for evidence to support applications for additional funding units promotes the need for a coherent system. Some project colleges needed to move from using a range of different tests across the college with no consistency of approach, administration, marking, interpretation or use of information to a whole-college approach underpinned by a clear strategy and policy, implemented through transparent systems and structures.

In most cases, something which started as small-scale activity with its roots in 'special needs', has become part of the learner's entitlement and is linked closely with a desire to offer appropriate support to a wide range of students in a bid to improve retention and achievement.

Where colleges can offer rigorous and robust initial assessment of learners it will be possible also to offer the individual support to ensure personal and academic progress. Data collected through rigorous initial assessment processes can provide a baseline against which to measure student achievement, to help motivate learning and celebrate success.

In short, there is a demonstrable value in making an early assessment of learners' needs so that difficulties do not grow into problems of non-attendance, missed deadlines, lack of progress/achievement and, eventually, drop-out. However, the outcomes of assessment must be used at different levels. Aggregate data need to inform strategic and curriculum planning and the allocation of resources. Information about individuals needs to be used positively to motivate learning, secure support and, where appropriate, additional funding units.

Purposes of initial assessment

It is important that colleges are clear about the purposes of initial assessment. Initial assessment can:

- guide placement
- identify individual needs and inform support provision
- inform strategic planning
- promote curriculum development and change
- provide evidence for additional funding units
- provide baseline data to motivate and provide a measure for student achievement.

Readers will need to consider the extent to which their own institution aims to do all of the above and the extent to which this is commonly understood by staff at all levels across the organisation.

Managing initial assessment in FE colleges

Overall management responsibility for initial assessment is divided equally across project colleges with 50% managed through client and student services and 50% through a senior curriculum manager. Although strengths and weaknesses are attributed to different approaches there is no evidence that the particular location of the service within the management structure has affected developments.

Usually the operational management of initial assessment is the responsibility of learning and study support managers and their teams. Colleagues with a range of related roles and responsibilities may be represented in teams which co-ordinate and implement initial assessment.

They include:

- learning and study centre manager
- learning and key skills workshop managers and staff
- English, Maths and IT staff
- library staff
- adult basic education staff
- literacy and language support staff
- Section 11 staff
- additional needs managers and staff
- managers and staff working with learners with difficulties and disabilities.

Learning support staff rarely implement initial assessment processes on a large scale. One favoured approach was supporting others in administering tests and/or developing and presenting initial assessment assignments. There was no consistent approach to marking, interpreting results and feeding back to students or others. Learning support staff, mainstream staff and tutors were all involved in different stages of the process. Outcomes of assessment were also collected, recorded and used in different ways.

One college saw initial assessment as the responsibility of threshold services. The diagnostic assessment officer, part of the threshold services team, serviced programme area teams so that initial assessment was programme-related, with coherence provided by a common policy and framework. Study Link then took up responsibility for support. Both were located in client services.

Overall, there are quite complex management models with clearly identified senior managers but potentially large numbers of staff with a wide range of specialisms and experience. Skills can be used to best effect where there is overt, real senior management support, a clear whole institutional strategy and policy and when roles and responsibilities are clearly described, understood and supported. A measure of the effectiveness of any management model must be the degree of consistency in the quality of the learners' experience.

Best practice

Best practice in the management and implementation of initial assessment includes:

- a common and shared understanding of the purpose(s) of initial assessment
- a clear management strategy which reflects the college mission and purpose
- an assessment policy and code of practice
- a transparent management structure which identifies roles and responsibilities, clear lines of communication and accountability.

In reviewing the management and implementation of initial assessment, colleges will need to reflect on the purposes of initial assessment and the extent to which existing policies, structures, roles and responsibilities will help the college achieve in relation to each of the identified purposes of initial assessment.

CHOOSING AND/OR DEVISING ASSESSMENT MATERIALS

In 50% of the colleges which worked with FEDA decisions about the kinds of assessment materials to be used and, where appropriate, their development, were made by specialist learning support staff and their colleagues in curriculum areas.

> *One of the most important aspects of this system has been the involvement of tutors. Their understanding of the issues, their knowledge of assessment measures and suitable approaches have been crucial to the success of systems.*

In the other 50%, decisions may have been made at senior management level or by specialist teams of guidance and admissions or learning support staff. Some colleges did not consider this good practice.

> *No time was allocated for liaison between support and mainstream staff. In a large multi-site college this led to difficulties: poor administration, inconsistencies in marking.*

Regardless of who was involved in decision-making all but two of FEDA's project colleges opted to use the Basic Skills Agency screening materials.

Other nationally available materials used to support initial assessment practice included:

- ACCUPLACER: an adaptive, computer-aided placement test developed for use in North American Community Colleges; tests maths and language skills at a range of levels (not diagnostic), chosen for general screening purposes
- AEB Achievement Test in Numeracy Level 3: put onto computer and chosen for use with adults
- Foundation Skills Assessment: a paper-based language and numeracy test which can be computer marked with an optical mark reader, chosen for use with adults
- MENO: Thinking Skills Assessments from the University of Cambridge Examinations Syndicate; Literacy, Spatial Operations and Understanding

Argument chosen for use with Access students, and Critical Thinking chosen for use with A-level students

- NFER-Nelson Basic Skills Numeracy Test: chosen by the learning support team in collaboration with numeracy tutors across the college, deals with calculations, approximations, problems
- NFER-Nelson General Ability Tests: chosen to present a profile of verbal, non-verbal, numerical spatial ability
- NFER-Nelson Graduate and Managerial Assessment: a battery of three high-level aptitude tests (verbal, numerical, and abstract reasoning) for use with adults aiming for higher education (HE.)

In all cases these tests were used with a relatively small proportion of the student population, often adults. In most cases they were introduced because colleagues were seeking to test learners on higher level programmes using appropriate and acceptable materials. Those who implemented tests needed to be trained to administer, mark and interpret results. These responsibilities did not constitute a significant aspect of staff roles and were not always formally recognised.

Where colleges decide to use a range of nationally available initial assessment materials or tests, it is critical that those with responsibility to administer tests are working in line with a Code of Good Practice similar to the draft in Appendix 1.

All colleges were involved in developing their own materials. Sometimes they were for initial screening but more often to be used as a final stage of initial assessment during the induction programme. An early task in the development of 'home-grown' assessment materials is the involvement of curriculum teams in identifying the skill demands of their particular programme.

One college developed its own key skills framework (skills fundamental to successful learning, not NCVQ's Key Skills) – see the example in Figure 1 (on page 13). With advice and guidance from specialist learning support staff, curriculum teams used the framework to analyse the skill demands of programmes before designing their own assessment materials to identify which skill demand could and could not be met by in-coming students.

Curriculum teams preferred this approach as they saw initial assessment as part of a continuum. They needed something which was related to the individual and their chosen programme of study, integrated and on-going within the programme, and critically – which would be seen in a positive light by students. Curriculum teams were also encouraged to use their analysis of the framework to identify how and where students could be helped to develop skills through:

- integration within the subject or unit
- discrete, taught sessions
- flexible learning
- learning/study support.

Another college (see the example in Figure 2 on page 16) used the expertise of its diagnostic assessment officer to draw up checklists, from which programme teams could produce a profile of course demands as a basis for developing programme specific screening materials.

Again, programme teams took responsibility for the development of materials, drawing on the expertise of specialist staff.

In a third college, an initial assessment and guidance manager had responsibility for co-ordinating all initial assessment, including the development of subject-specific language materials to help the placement of students without the formal qualifications identified in course entry criteria. Staff perceptions were identified as an important issue as, for example, it became clear that 'a lower standard for first language speakers appeared to be used in Art and Design than other courses despite the fact that the language level is high as well as wide ranging'.

> *The Art and Design test was devised particularly to see if it might give more exact information about course-related language skills. For example, the reading text was based on a design brief and aimed to discover students' abilities to cope with basics: descriptive vocabulary and the language of basic design concepts. The writing task was primarily descriptive because an ability to describe visual experience was identified by Art and Design staff as a basic criterion for the course.*

This serves to reinforce the need to be clear about the skill demands of each learning programme and for specialist and programme staff to work together to develop, or support the development of materials.

Figure 1 An extract from a college-devised skills framework used to profile course demands

Information processing	Level 1	Level 2	Level 3
Sorting, classifying and organising information	One set of materials against a few given headings	From a variety of sources Into main and sub categories	With cross referencing Generate rules and use to predict
Drawing conclusions	From given information on a single topic	From a variety of information under some direction, e.g. for given assignment Draw relevant conclusions from new data	From a variety of information unaided Show awareness of limitations
Summarising	Identify main points of paragraph or short extract	Summarise whole article, talk or long extract	Summarise and synthesise information from a variety of sources
Argument	Explain and justify own choices, e.g. of materials or treatment	Identify/express arguments in support of a proposition Adduce evidence	Analyse arguments and use to refute a proposition
Comparing and contrasting	Compare/contrast two or three items under direction	Compare/contrast two or three items	Compare/contrast any number of items
Analysing	—	Break simple material into component parts	Break complex material into component parts Identify appropriate other response
Evaluating	One topic against a few given criteria	One topic against a range of given criteria Several topics against a few given criteria	A range of topics against self-determined criteria

Figure 1 An extract from a college-devised skills framework used to profile course demands continued

Organisation	Level 1	Level 2	Level 3
Paper management	Keep notes and records sorted by topic or module	Set up and maintain a file system with main, sub headings and index	Set up and maintain file system which includes cross-referencing
Time management	Attend on time Meet deadlines for a sequence of tasks	Manage free study time appropriately Manage concurrent tasks and meet deadlines	Prioritise concurrent tasks Plan ahead
Revising	—	Undertake revision where necessary	Plan and implement a revision timetable
Self management	Co-operate with tutor to identify strengths/ weaknesses Agree short-term goals Review progress	Plan work to improve on weaknesses With support: evaluate own performance, set medium-term goals, review progress	Set own short-, medium- and long-term goals Review own progress Evaluate own performance
Information gathering			
Information seeking	—	Use library and flexible learning centres with support Identify and use external sources of information with direction Design and undertake a simple survey	Use library and flexible learning centres unaided Identify and use external sources of information freely Design, pilot, modify, undertake and evaluate a simple research tasks
Note taking	Simple notes in pre-set formats, e.g. gapped handouts	From text or talk with support, e.g. on a familiar topic	From text, video, talk or audio material See also Language: Extracting key information, detailed understanding, discussion skills, inference

BEST COPY AVAILABLE

Figure 1 An extract from a college-devised skills framework used to profile course demands continued

Information output	Level 1	Level 2	Level 3
Understanding tasks or questions	Follow simple written or verbal instructions with a memory aid	Analyse components of task Make choices Follow instructions without memory aid	Complete tasks from outline instructions only Identify not only content but also appropriate tone / treatment
Use appropriate written / oral / graphic formats	Brief written or verbal reports Forms e.g. accident reports Multiple choice/short answers	Semi formal report Small group presentation Formal letters Essays	Formal / lengthy reports, e.g. extended essay, technical report Individual oral presentation
Planning and drafting	Planning stage in some	Planning stage in all work First draft stage for major course work assignments	Second draft stage in major piece of work
Proof reading	Check content for accuracy Proof read for spelling, grammar and punctuation with support	Systematically proof read for spelling, grammar and punctuation Check content for accuracy and tone	—
Handling tests and exams	Follow instructions correctly	Manage time within test or exam Practise beforehand	Strategies for coping with strengths and weaknesses *See also Language. register, tone, style, paragraphing, sentence structure*

Figure 2 An example of a college-devised checklist identifying entry requirements

Styles of writing needed	Essential	Taught	Not relevant
What sort of writing are students required to do?			
• Copy notes • Notes for own use • Supply single word answers • Make short answers on factual matters • Write a descriptive paragraph of factual matter • Descriptive writing with explanatory commentary on content • Descriptive writing with critical or evaluative commentary • Persuasive writing • Analytical writing • Imaginative writing • Selection and use of type of language – register – e.g. informal/formal – level of language – technical • Knowledge of particular layouts: – letter – report – essay • Other			
Writing skills checklist			
Content:			
Writing is relevant to task set Sufficient detail has been included Irrelevant information is excluded Ideas flow logically Ideas are clearly expressed Ideas develop an argument or analysis			
Language:			
Everyday appropriate vocabulary Specialist vocabulary is used Style of language is appropriate			
Conventions:			
Clear presentation Grammar acceptable Spelling acceptable Punctuation acceptable Formal writing structure is used: introduction, body, conclusion			

Choosing/devising initial assessment materials

In choosing or devising initial assessment materials, colleges need to:

- be clear about the purpose for which tests will be used
- identify the skill demands of particular programmes and ensure that initial assessment materials relate to them
- ensure that specialist learning support staff can be available to advise, guide or work with curriculum teams
- clarify roles and responsibilities in line with college policy and a code of practice.

TIMING INITIAL ASSESSMENT

The stage at which initial assessment is introduced to learners relates to purpose. Where the purpose is to inform judgements about the type or level of programme, the learner needs to be assessed pre-entry. This may be through an interview which seeks evidence of a match between previous experiences and skills and those demanded by the level and kind of programme. Qualifications are often recognised as evidence of previous experience and skills, and where students do not have the qualifications stated in the formal entry requirements some colleges have chosen to implement pre-entry testing. Assessment is wider than testing and Chapter 3 seeks to clarify technical aspects of both.

> *Students are only tested on the main college test if they do not have the required minimum GCSE or (G)NVQ or equivalent qualifications from a reputable institution. At present students should be tested if they do not have one of the following:*
>
> - *D in GCSE English (E for GNVQ level 2)*
> - *Communication at GNVQ level 2*
> - *other nationally recognised equivalent.*

In this example, initial assessment is focused on language skills and the college uses the tests to determine whether applicants will be able to deal with the course for which they are applying. Where a college has an open access policy such testing needs sensitivity. It must be made clear to learners that the outcome of testing will inform the level of programme which the college believes will best suit the learner and through which they will be most able to achieve.

Some colleges screen for levels of basic skills, pre-entry, to identify the learners most likely to benefit from learning support on-programme.

Figure 3 An extract from a college's Initial assessment guidance for tutors

Q Is the screening about getting people onto the right courses?

A The screening exercise is not a screening out test. It shouldn't be used for pre-entry selection. The exercise tells us nothing about the nature of the particular difficulty any student may be facing. It also says nothing about their potential to develop over the duration of the course. Getting students on to the right course is more rightly the concern of the initial interview and induction.

This exercise is to be used after the student has started on the course. It tells us where a student stands in broad literacy and numeracy terms in relation to the literacy and numeracy levels required to complete the course.

However, some students do end up on inappropriate courses. With sensitivity and discretion the screening results might form part of the evidence which suggests that students should be re-routed onto a more appropriate course – be it at a more or a less advanced level.

If the outcome of testing is likely to inform selection, or to lead to re-routing where a choice has been made and a place offered, this needs to be made clear to learners.

Where testing is used pre-entry for a large number of learners, the information generated can be used at whole-college level to inform curriculum planning and resourcing as well as support services. However, the benefits of being able to plan early need to be balanced against the costs of pre-entry testing. The most significant costs centre around students who don't enrol after testing – some because, regardless of testing, they choose to go elsewhere and others perhaps according to college anecdotal evidence because they were adversely affected by testing. It is useful to compare numbers of applications with enrolments. Large scale use of pre-entry assessment is not cost effective if the conversion rate is not good.

Most colleges did not screen pre-entry, as they did not want the students' perception of testing to present a barrier to access.

Screening learners at entry is now common practice in most tertiary and general FE colleges. The purpose of screening at this stage is to identify students who need support. However, in some colleges students who have been accepted onto programmes at particular levels have been re-routed where screening suggests that there may be difficulties. Colleges need to be clear in communicating to students the purpose of screening and the possible outcomes. If significant numbers of students are re-routed after testing it may be necessary to re-evaluate the college's admissions/selection procedures.

> *Initial assessment is part of the continuum of gathering information about students that begins with recruitment. It is a post-enrolment activity that contributes to the development of individual learning programmes. Its place in the learner pathway is clearly identified in threshold/entry services.*

Although the Basic Skills Agency's (BSA) screening tests are the most commonly used, several colleges have developed their own screening and diagnostic assessment materials. Both Basingstoke and Sheffield Colleges' materials are bought by other colleges across the sector. Screening tests used by colleges to assess students at entry are limited in scope, and focus largely on literacy and numeracy skills. BSA screening tests and a very limited range of commercially produced materials are used widely with students wishing to follow programmes at level 2 and below. Where colleges screen students on higher level programmes, they frequently use screening materials developed by themselves or other colleges.

Screening at entry identifies an approximate level of skill. Learners who are more than a level below that required to function on their chosen learning programme will be targeted for support. Experience suggests that students need to be identified and programmed for support at the earliest possible opportunity:

> *Students in need were arriving for support very late in their course, so the help we could give them was less effective than it could be. Typically, they were arriving having already under-achieved on assignments or unit tests, and tended to be demoralised and negative.*

The guidelines prepared for tutors administering screening tests in one college signal the importance of screening early; give a deadline for the return of test results; but at the same time offer some flexibility to programme teams about exactly when testing takes place.

Figure 4 An extract from guidance on assessment

Screening should take place as early as possible. It obviously helps us if we can get a basic skills profile of our students quickly. It also benefits those students needing support to be identified as soon as possible. However, the exercise can be carried out at any time during the first three weeks of the course. This will mean that course teams can choose the time and setting most appropriate to their students. Regardless of when a course team decides to complete the exercise, we need the results back no later than Friday 30 September.

Feedback from adult students on an Access programme in another college indicates that the MENO Thinking Skills assessments were introduced too late in the course:

> *I feel that the MENO booklets should have been given out at the beginning of the term and not in the middle of assignments.*

The outcome of a screening test signals whether a student may need support but does not provide a detailed profile of what a learner can or cannot do, some colleges have developed diagnostic assessment materials so that further judgements can be made through the induction process. Where there is good practice, tools will be rigorous and robust and will have been developed in collaboration by vocational and specialist staff, in the light of a programme specific skills profile. Only students who have been flagged up as needing support will require assessment at this stage.

Again, it is important to administer, mark and feed back to students as soon as possible.

Initial assessment and feedback

Initial assessment, together with student feedback and negotiating support, should usually take place during the first four weeks of a programme. This helps to avoid confusion with 'selection', and also ensures group and individual needs are identified early. The exact timing will depend upon the nature and length of a particular programme.

Unlike the screening tests, which are quick and easy to mark, diagnostic assessments can be very time consuming. One college reported three to four hours' marking for one diagnostic assignment for a tutor group of 20 students.

Even a sophisticated initial assessment process may miss students who have support needs. It is important to remember that assessment is an on-going process and if there is evidence that a student is failing to make progress, learning support may help.

The outcomes of screening will need to be available early enough to identify which students will benefit from more detailed assessment through the induction process. A clearly communicated schedule which identifies what will happen and when, with dates and deadlines, will smooth the process of initial assessment across the college.

Purpose and timing of assessment

The purpose and timing of assessment need to be considered so that:

- assessment which seeks to aid placement takes place pre-entry
- assessment which seeks to identify skill levels and support takes place at entry
- assessment which seeks to identify the specific nature of support needs is an integral part of the induction process for those students who have come through screening as needing support

ADMINISTERING INITIAL ASSESSMENT

In this context administration has been interpreted as introducing the assessment process to students, giving out materials, supervising and, where necessary, timing students' engagement with tasks and collecting in papers. Most colleges chose to administer (but not always mark) the Basic Skills Agency screening tests through programme teams. A few colleges chose to assess applicants pre-entry and used specialist guidance or learning support staff.

A critical part of the process is the way in which staff have been prepared for their role. They need to be clear about the purpose of the initial assessment and to understand the need to communicate this information to the students. Some colleges offered staff briefing sessions for programme teams and some produced written guidelines.

Figure 5 An extract from a college's guidance on initial assessment produced for tutors

Therefore, it is important that we think about how the exercise is presented to the students and that we are clear as to why we are asking them to complete it. As such, it is worth letting the students know that:

- there is no pass or fail mark.
- no one group is being singled out.

All full-time students are being asked to complete the exercise from Foundation to Access and A-level.

- the screening exercise will help us plan provision more effectively. In that sense, the exercise is essentially about our teaching meeting their needs
- the results are confidential.

Students need to feel reassured that the outcome of their screening tests will be fed back to them and will be used for their benefit and the benefit of the college.

Confidentiality is an issue. Clearly individual results will not be shared with other students but in most if not all cases they are shared with other staff. In psychometric tests, raw results remain confidential to the tester but key messages or trends may be communicated to other staff, as well as being explained to the student concerned. Colleges need to be clear in communicating to students who will have access to data, in what form, and why.

Where colleges developed their own initial assessment materials, programme teams usually administered the range of tests, apart from, in some colleges, the writing tests. Again staff briefing, guidelines and support were given a high priority as was scheduling (see Figure 6 on page 20).

Figure 6 An extract from a college schedule for managing initial assessment: delivery of assessment

Who delivered	When delivered	Staff preparation	Student preparation	Timescale	Organisation
A Course leaders to all students	Week 1	Experience from previous year Project team support	Preliminary explanation of rationale for assignment to whole group Detailed explanation with personal tutors in three groups	By end of week 1	Students allocated at random to three groups
B Course leader with course leader for Advanced	Week 1	Briefing by course leader Project team support	Self-assessment checklist completed prior to assignment in class Student alerted to assignment function as assessing key skills	By end of week 1	Joint induction with advanced students (35) groups of 4/5 Completed independently Hand-written not word-processed
C Range of lecturing staff	Week 1	None other than course leader Project team support	Cover sheet explaining rationale Course leader gave explanation to whole group	By end of week 1	Students chose own groups
D Course leader	Week 2	Project team support All staff briefed to be ready to help students	Given specimen worked assignment from Business Studies to look at General explanation but no specific reference to assessment	By end of week 2	Groups of 6 Screening tasks to be completed individually

One college experienced problems where programme teams were not aware of the total picture; were not supplied with all the necessary materials; and were unclear about how the outcomes would be fed back to students and used by the college. Briefings for further assessment were planned around the following checklist:

- The tests should include a reading comprehension and a piece of continuous writing.
- Reading comprehension should be based on a piece of text which students will have to read as part of their course if it is a course-specific test.
- Clear written instructions should be given to candidates.
- There should be a written marking scheme which indicates:
  - acceptable level
  - acceptable level with additional support
  - skills not good enough
  - skills too good – refer to higher level
- There should be criteria underpinning the marking scheme which are made explicit to candidate (e.g. 'You will be marked on the accuracy of your English, your ability to organise your material and the level of your vocabulary and sentence structure').
- A feedback sheet identifying any support requirements should be given to the student.
- There should be a sheet to go to client services indicating that a need for additional support has been identified and any possible information and nature of need.

Evaluating early experience of administering programme-specific assignments through the programme team's induction process, student feedback in one college indicated:

- students felt comfortable with the assessment process, but were not always clear about the purpose and the link to learning support
- the materials were not always seen as relevant to the programme of study
- students were unclear what would be judged and how performance would be measured
- over a third of students found the tasks 'too easy'.

Planned improvements included:

- The content and the specific skills assessed should be clearly relevant to the course and this relevance should be explained to the students.
- The assignment should be demanding enough to give information on students' competence at a number of levels.
- The skills assessed should be, not only basic literacy and numeracy, but also organisation, personal and study skills.
- Students should be more closely involved in the process of assessment:
  - they should know that they are being assessed and what criteria are being used
  - they should have detailed individual feedback on their strengths and weaknesses
  - they may even be involved in grading decisions on their own or peers' performance
  - they need to be given more information about the standards expected of them.

Both staff and students need to understand clearly the purpose of the initial assessment exercise. Materials need to be seen to be relevant to the purpose and will be better received if students perceive them to be relevant to the chosen area of study. There is evidence that students will perform better when they are clear about what is being measured and how judgements will be made. Both staff and students need to know about time, and any other constraints.

Psychometric tests in the project colleges were always administered by specialist staff in line with codes of practice. However, other materials bought in for initial assessment were sometimes introduced by programme teams. For example, the MENO Thinking Skills assignments were introduced to students on an Access course by members of the core skills team, supported by specialist learning support staff, and students were able to complete quite extensive pieces of work in their own time.

Much care went into the introduction of externally produced assessment materials and students were often given opportunities to practise before engaging with a formal assessment piece. Where colleges have specialist staff administering national tests to small, discrete groups of students, it is important to draw on their experience so that, where appropriate, it informs the college's overall approach to the implementation of initial assessment.

Figure 7 An example of instructions to students in a college-devised task to assess writing skills.

Using the tasks

Students must be told why and how they are being screened. For example:

You are to be asked to do some writing to see how you:

- *write to the point*
- *express yourself clearly*
- *plan/organise your writing*
- *present your work*
- *use grammar, punctuation and spelling.*

Make sure the instructions are clear. For example:

- *You will have 30 minutes to answer the question.*
- *Spend the first five to ten minutes reading the question and working out how you're going to answer it.*
- *You should aim to write about 200 to 300 words – about a side to a side and a half of A4.*
- *Concentrate on getting your ideas across clearly.*

To manage the process well, colleges need to be clear about the roles and responsibilities of all concerned. It may be helpful to start from an ideal student experience and work back from there in making decisions about who needs to perform what tasks and how they need to be briefed and supported.

Administering initial assessment

Colleges need to develop:

- clear implementation strategies which relate to the purpose of initial assessment, in line with the college's assessment policy
- clarify and confirm the roles and responsibilities of all staff involved in the process
- establish clear and simple systems for managing the process of initial assessment
- offer training, guidance and support to staff to ensure they feel comfortable and confident in their role
- communicate clearly to students the purposes and processes of initial assessment.

MARKING AND FEEDBACK

Marking Basic Skills Agency screening tests and 'home-grown' initial assessment materials was sometimes done by programme teams and sometimes by specialist learning support teams. The former approach can spread the load and should make it possible to mark large numbers of tests in a relatively short time, to provide immediate feedback to students and set up support before students become demotivated by difficulties.

In fact, some colleges found that marking was not done immediately, took more time than anticipated, and occasionally had to be redone because of inaccuracies. One college noted that 'Marking was extremely variable and must significantly affect the usefulness of the tests.'

Positive outcomes from involving programme teams in marking initial assessment exercises include:

- the raised profile of basic skills and basic skills needs with large numbers of teaching staff
- a recognition of the need to change classroom practice through more inclusive approaches to teaching and learning.

These benefits must be worth working for and where marking needs to be improved, staff training, guidance and support will improve the accuracy and consistency across the college. See the example in Figure 9 on page 34 of staff guidance for marking a writing task.

Figure 8 A college schedule for marking and feedback from initial assessment

Who marked	Marked by when	How marked	Who gave feedback	When fed back	How fed back	Comments
A Personal tutors but not own tutor group	Tutorials marked in week 1: delays thereafter due to staff illness	Answer sheet for Maths Language support tutor devised criteria for writing Course leader adapted into tick list Took about 10 minutes per assignment	Personal tutors to tutor groups	In first tutorial in week 1	Individual interview	
B Course Leaders	Week 1	First trawl 'gut reaction' Using checklist Answer sheet for Maths	Course leaders	Week 1	Individual interview as part of action planning	Checklist cut time in marking Some students reluctant to ack. need Setting ass. in course context helped
C Course Leader	Week 3	Course leader prepared own Maths marking guide Language support tutor prepared criteria for writing Simplified by course leader for use	Course leader	Week 4 Weakest students given feedback first	Individual interview Very casual in order not to provoke negative atmosphere	
D Course Leader	Week 5	Initial trawl by 'gut reaction' Then own checklist	Course leader	Week 6	Ind. written feedback on assignment front sheet General verbal feedback to whole group Invitation to discuss if wanted	Students who used graphics package avoided having to do the maths calculations No record of poor performers

Figure 9 An example of staff guidance on marking a writing task

The aim is to build a profile of the skills and knowledge a student has demonstrated. The skills usually fall into three categories:

- content, e.g. relevant use of language
- conventions, e.g. punctuation, spelling, grammar
- presentation, e.g. handwriting or layout.

If the writing task is short, these can be assessed in one reading. More complex tasks may require two readings, one for content and then a check on conventions.

Depending on the level of the programme, the levels of language, spelling and punctuation will vary. It is impossible to give precise guidelines but the table below offers some guidance to the levels which should be expected at entry.

Level	Language	Grammar and structure	Punctuation	Spelling
1 Foundation	Simple	Sentences	Capital letters	Most everyday foundation words
2 Intermediate	Simple, clear	Sentences	Full stops, capital letters	Everyday words
3 Advanced	Basic adult level	Sentences, paragraphs	Full stops, commas, capital letters	All except unusual words
4	Wide ranging	Introduction, main body, conclusion	All common including apostrophes, colons	All words

Within each of those categories a margin of error may be acceptable e.g. two or three minor errors may be acceptable.

The assessments can be recorded on the screening grids which provide individual and group profiles. A tick or a cross is made under each of the skills/knowledge headed columns against students' names. In practice there are four assessments possible:

- student's performance at an acceptable standard
- student's performance not at an acceptable standard
- conflicting or uncertain evidence – further monitoring and discussion needed
- no evidence available, further monitoring and discussion needed.

Individual student profiles across the range of entry criteria are read across the page and group needs for each criteria can be read down the page.

Students with crosses in the contents columns who cannot write relevant ideas or who are difficult to understand, should be referred for individual support. The conventions of spelling, grammar, punctuation and presentation may result in a referral if the problem is severe and if these are individual and not group needs.

BEST COPY AVAILABLE

Where programme teams are able to work with their students to identify their difficulties early they are more likely to want to share and 'own' problems and to help provide solutions.

Some colleges introduced quite complex models for marking student work produced through an initial assessment process, and feeding back the outcomes to learners and to appropriate staff. Simple, transparent systems work best and colleges keen to improve their practice will do well to remember the purpose of the exercise is to target appropriate support at students who need it, before difficulties precipitate themselves as problems of non-attendance and drop-out.

Where programme teams find a significant proportion of new intake are assessed as needing support, there may be an argument for re-examining and re-defining entry criteria. Learning support can not compensate for a poor match between the experience and skills of the learner and the demands of the learning programme.

All students need to be told how they have been judged through the initial assessment process. Feedback needs to be given as soon as possible after the assessment process. Colleges which evaluated student perceptions of the initial assessment experience found that 66% of students reported feelings of low self-esteem after testing, with 70% making positive comments about themselves after feedback.

Programme teams need to have access to aggregate data for groups they teach. They will need access to a clear and simple system for communicating information about both individual needs and group profiles to learning support managers.

Information should be used to plan additional support for individuals where necessary and appropriate as well as to inform strategic planning.

Students and all the staff who teach and support them must have access to a simple system which recognises, monitors and tracks progress.

Marking and feedback

Colleges need to:

- ensure consistency in the quality of marking of initial assessment tasks and tests
- have a clear policy on confidentiality
- provide immediate feedback to learners
- communicate appropriate information from the outcomes of the assessment process to:
  - tutors
  - programme tutors
  - managers.

3 Technical aspects of initial assessment

Assessment and testing

Assessment is not synonymous with testing. Objective tests provide one means of obtaining information about students – in particular about their potential for success and their support needs. While most people feel they know the difference between 'tests' and general forms of assessment, it is very difficult to define what 'objective tests' are. Any attempt to provide a precise definition of a 'test' or of 'testing' as a process, is likely to fail as it will tend to exclude some procedures which should be included and include others which should be excluded.

Within the context of FE initial assessment, objective tests are procedures which are used to make inferences about a person's ability to cope with the demands of various programmes and, by implication, their likely support needs. As such, they are used as forward-looking assessment tools. Tests of ability and aptitude are forms of assessment concerned not with assessing what you can do now or what you have achieved in the past, but with making inferences about your potential to achieve in the future.

Tests are those assessment methods which:

- provide quantitative measures of performance
- involve the drawing of inferences, for example, about a person's potential to learn or the reasons for them experiencing difficulties in learning, from samples of their behaviour
- can have their reliability and validity quantified
- are normally designed to be administered under carefully controlled or standardised conditions
- embody systematic scoring protocols.

Any procedures used for 'testing' in the above sense, should be regarded as an 'objective test'. All such tests should be supported by evidence that the assessment procedures they use are both reliable and valid for their intended purpose. Evidence should be provided to support the inferences which may be drawn from the scores on the test. This evidence should be accessible to the test user and available for independent scrutiny and evaluation.

Much of what follows will also apply to assessment procedures which lie outside the domain of 'tests', interviews, appraisals of records of achievement, and so on. Any assessment process which, if misused, may cause personal harm or distress should be used carefully and professionally. Misdiagnosing a person's learning support needs may not only waste college resources, but damage the self-esteem and confidence of the individual concerned.

Objective tests differ from other forms of assessment in that they are based on a technology which makes it possible to specify what they are measuring (their validity) and how accurately (their reliability). Typically, educational assessments (such as tests of college attainment, GCSEs and other school examinations) do not have these qualities: their validity is a matter of judgement and their reliability often unknown or unassessed.

Objective tests assess a broad range of human characteristics: personality, motivation, values, interests, ability and achievement. However the present guidance is only concerned with one subset of this complex area: the assessment of basic or key skills and the diagnosis of specific learning needs. (For more detailed information, see Appendix 2.)

While objective tests can be used as a mechanism for making judgements about learners, it is important to remember that assessment is more than just testing and that other activities can be effective in providing evidence of students' knowledge, skills, attitudes, performance and needs. Indeed, FEFC guidance on completing the Additional Support Costs form assumes that evidence for claims will come from a range of assessment activities:

> *Institutions will use a range of assessment instruments and strategies throughout the learning programme to identify students' additional learning support needs. The assessment carried out should be relevant and identify an individual's need within the context of the curriculum followed.*
>
> *(FEFC 1996-97)*

This broader level of thinking needs to be encapsulated in the college's assessment policy which needs to be known and understood by staff.

Assessment policy

Managing assessment (FEDA 1995) examines the development of a whole college strategy for consistency and quality in the management of assessment and sets out key features of an assessment policy. Elements to be included in an assessment policy are:

- assessment principles
- assessment stages
- assessment processes.

Key principles of assessment apply to all stages in the learner pathway and to the vast spectrum of programmes with different assessment regimes. These principles need to underpin initial assessment practice for all learners:

- enhancement of learning: a key purpose of assessment is to ensure learning
- reliability and validity: all assessment should be based upon explicit assessment objectives
- shared understanding of standards: staff are trained in assessment
- quality assurance: a system is in place to monitor assessment practice.

(FEDA, 1995)

Consistency of initial assessment practice should be guided by a whole college policy, with student entitlement laid down in the student charter. It may be helpful to consider extracts from one college policy, (see Figure 10) to compare this with your own.

You may want to reflect on your college's initial assessment practice and consider the extent to which it reflects general principles of fair assessment, the excerpts above and your own college policy. Such reflective practice will be instructive as colleges move to become more self-critical as a means of improving the quality of all aspects of college services and provision.

Figure 10 An example of a general assessment policy statement adapted from a college's student charter

Assessment is an integral part of all learning activities at X College. It is natural and relevant to all students and enables both the learner and the tutor to identify learning that has taken place, plan the next stages, motivate further learning, and encourage development progress.

Assessment regimes and procedures for the programme as a whole will be explained, at an appropriate time, to students by members of their course or programme team. In addition, students will be given information on individual assessment, to include:

- what is assessed
- the criteria for success
- when to be completed
- when it will be marked and returned
- information about results and performance.

Initial assessment: clarification and definitions

Initial assessment is the earliest assessment of learners as they move through the pre-entry, and entry stages. Initial assessment processes need to reflect the spirit of the college's policy statement and the entitlement set out in the student charter.

In most colleges initial assessment is a staged process and probably starts with pre-entry guidance. Here the learner and the guidance specialist look together at previous experience, 'at learning that has taken place' with the aim of identifying the most appropriate choice of learning programme so that personal and academic progress is made.

Judgements made here will be likely to be conditioned by the learner's awareness of self and their ability to articulate that self-knowledge along with their aspirations. Documentation which the learner may bring, like a careers action plan, record of achievement or portfolio of work can all inform discussions and decisions which are made about learning pathways.

In certain situations potential students may be assessed through the use of tests, for example, aptitude tests, basic skills tests. The purpose of testing at this stage should be to help the learner form a more accurate picture of themselves so that they are better able to consider information and advice about learning programmes in relation to individual needs and aspirations. The purpose of testing should be communicated to students.

SELECTION: PLACEMENT OR CLASSIFICATION

Selection is a complex process. It involves all that happens from the time a potential student approaches the college, possibly only seeking information, to when they start their learning programme. Selection involves two processes, placement and classification. There can be tensions between these as the first focuses on the needs of the student applicant, while the second concerns the resources of the organisation to which they are applying.

Placement means placing the learner on the programme best suited to them after considering prior experience, interests, skills, and future plans, alongside detailed information of programmes. Judgements about learners will be conditioned by their awareness of self and evidence of their achievements.

Tests can be used in this context to get a measure of general aptitude or ability or perhaps to get a measure of competence in basic skills. In this context, a learner who demonstrated a low level of numeracy skills would not be placed on a programme which required high-level numeracy skills. Effective placement decisions are best made by those who do not have a vested interest in recruitment or growth targets so that learners can benefit from disinterested advice.

Classification, on the other hand, involves optimising the assignment of applicants to places available on programmes within the college. Classification is a process of finding the best overall fit between applicants and places. In this context, 'best' means the most cost-effective in terms of organisational outcome.

In practice, applicants will select a programme through admissions procedures and will be advised in relation to the demands which it will make on them. If there is perceived to be a good match between the applicant's interests and capabilities and the demands of the programme and there are places available, they will be likely to be selected for it. FEDA strongly supports placement as good practice in admissions procedures and does not wish to promote the process of classification at the expense of the student's individual best interests. We do, however, recognise that organisational constraints will affect the opportunities available.

THE SELECTION INTERVIEW

The assessment process which informs selection is usually in the form of an interview. Interviews will be most effective when the interviewer is clear about both the key requirements of the programme and the kinds of evidence which can demonstrate that the interviewee can meet them. Objective tests can contribute to that evidence. In particular, they can show when applicants have a potential for success which is greater than their achievements to date would indicate.

Earlier FEU work on entry criteria and GNVQs in 1995 suggests over-reliance on evidence of prior achievement in admissions interviews, rather than exploration of future potential.

The following tables provide information on the dimensions of achievement which colleges considered important, and the sources of evidence on which they rely. A group of colleges identified the dimensions of achievement which they believed were important to student success in GNVQ programmes. They also identified the sources of evidence which they were likely to note as indicators of student achievement. These indicators and dimensions formed the basis of a national survey on entry criteria and GNVQ. Table A indicates respondents' judgements of the importance of the various dimensions. Table B shows the rank order, drawing from the percentage of colleges indicating each source of evidence as being important (colleges could, of course, indicate more than one). Finally, Table C (best indicators for each dimension) shows which sources of evidence were considered to be the best indicators of each of the dimensions of achievement.

This illustrates that at the time of the survey there was a clear dependence on GCSE grade profiles as best indicators of some of the dimensions likely to lead to success on GNVQ programmes. Performance in GCSE was the source of evidence most used as a basis for matching students to programmes. However, GCSE performance was not seen as a good enough indicator of learning support need. At the same time, survey colleges were also introducing screening and diagnostic assessment to identify learning support needs.

Table A Dimensions of achievement identified as important by colleges

Dimension	%
Academic ability	46
Motivation/interest	25
Ability to communicate	22
Vocational knowledge/skills	18
Potential in vocational programme	16
Organisational ability	13
Relevant work experience	10
Maturity	10
Flexibility in learning	10

Table B Indicators of student achievement in rank order of percentage of colleges identifying each

Source of evidence	%
4–5 GCSEs Grades A–C	31
References/reports	27
National Record of Achievement	25
Portfolio of evidence	24
Previous GNVQ achievement	23
4–5 GCSEs Grade D–E	22
BTEC First	21
NVQs Level 2	21
GCSE Maths Grade A–C	18
4–5 GCSEs Grade below D–E	17
GCSE English Language Grade A–C	15
Other (e.g. interview performance)	12
Core skills attainment	11
Specific tests	3

Table C Perceived 'best' indicators for each dimension

Dimension	Indicator	%
Academic ability	4–5 GCSEs Grades A–C	72
Vocational knowledge/skills	References/reports	35
Motivation/interest	4–5 GCSEs Grades A–C	47
Relevant work experience	Portfolio of evidence	22
Ability to communicate	4–5 GCSEs Grades A–C	42
Flexibility in learning	4–5 GCSEs Grades A–C	22
Organisational ability	Portfolio of evidence	25
Potential in vocational programme	4–5 GCSEs Grades A–C	30
Maturity	4–5 GCSEs Grades A–C / References/reports	19

Initial assessment which informs placement processes needs to identify the best possible match between the learner, and the level and kind of programme to be followed. GCSEs may well constitute evidence of some dimensions of achievement but there is an argument for seeking a wider range of evidence. As a general rule evidence of achievement in the past is a good indicator of future potential. However, lack of past achievement does not necessarily indicate poor future potential. Past achievement is an outcome of a complex mix of personal and situational factors (ability, opportunity, motivation, the learning environment and so on). As colleges are potentially able to provide opportunity and supportive learning environments, those with ability who lack a record of prior achievement may succeed in the future.

Screening and diagnosis

The earlier section of this report confirms that many colleges use tests for both screening and diagnosis. It is important to understand the difference between these two related processes. One way of thinking about the differences between screening and diagnosis, is to imagine looking at a scene through a camera with a zoom lens.

With a wide-angle shot, you get a good overall impression of what is in the scene, what sort of scene it is – but you do not get any of the details.

By zooming in on some particular part of the picture, you get a lot more detail about that part – but lose sight of the overall picture.

Screening tests give you the wide-angle shots. Diagnostic tests provide a more detailed set of pictures. It is a matter of ensuring that the test chosen is best suited to the purpose for which it is being used.

Another way in which they differ is that diagnostic testing is about individuals, screening is about groups or populations. We 'screen' groups of people to identify those who are likely to have some particular quality, strength, or problem. We use diagnostic tests to identify the nature of an individual person's qualities, their strengths and weaknesses.

Psychologists use test batteries like the Weschler Adult Intelligence Scales (WAIS), the Stanford-Binet and the British Ability Scales for in-depth individual diagnostic testing. As well as giving the wide-angle view of an individual's ability, they can be used to focus on specific areas and aspects of ability to give a much finer-grain analysis. These individual diagnostic batteries require a background of experience in psychological assessment, and extensive training and skill to use properly. In practice, people requiring this level of diagnostic testing should be referred to a suitably qualified chartered psychologist for assessment.

On the other hand, tests like the NFER-NELSON Basic Skills Tests, or the Psychological Corporation's Foundation Skills Assessment (FSA), which were referred to earlier (see page 12) are more widely available to suitably qualified users. These are designed for use as group tests – providing the broad, wide-angle view. They are not intended to provide a very detailed diagnostic breakdown, at the individual level, of areas of strength and weakness.

The NFER-NELSON Basic Skills Tests measure both basic literacy and numeracy skills, and are designed for use with adults who have few, if any, academic qualifications. The literacy test is based around a newspaper from an imaginary town, while the numeracy test assesses the ability to carry out simple calculations, estimation and application of numerical concepts to everyday problems.

The Psychological Corporation's Foundation Skills Assessment is also designed to provide measures of attainment in basic numeracy and literacy skills for adults using materials which relate to everyday situations. The FSA consists of four tests, covering Vocabulary, Reading Comprehension, Number Operations and Problem-Solving, each available at three levels of difficulty (A, B, and C). There is a short screening test available to indicate which level of the FSA is most appropriate.

There are four possible outcomes from any screening process. Where a test is being used to indicate whether or not people might need learning support, these can be characterised as follows:

	Person does not need support	**Person does need support**
Test outcome 'positive'	**A** false alarm or false positive	**B** hit
Test outcome 'negative'	**C** correct rejection	**D** miss or false negative

BEST COPY AVAILABLE

In other words, a screening test will sometimes say that a person needs support when they do not (a 'false positive' result) and sometimes will fail to detect (a 'miss') someone who does need support. The number or proportion of positive outcomes can be varied by changing the test's cut-off score. Most screening tests for basic or key skills are designed so that the lower the score, the more likely the person is to need support. So, the cut-off score is the score below which a person's result is classed as 'positive', and above which it is classed as 'negative'. If you move the cut-off higher up, you would get more positives; move it further down and you get fewer positives.

For any particular cut-off score, the reliability of the test will determine the consistency with which people are classified, and the validity of the test will determine the proportion of people whom the test has correctly classified on any given occasion (i.e. those in cells B and C in the above table as opposed to those in A and D). These two points are very important. If a test is unreliable, then sometimes a person will 'pass' the cut-off score and on other occasions the same person will not. All assessments have a margin of error. For objective tests, the size of this error is known from the test's reliability. As a result we can make allowances for the risk of misclassification due to error in the measurement procedures. This is something we cannot do when we use assessment methods of unknown reliability.

Validity is different. It is about what the test measures, not how accurate it is. When a test is not valid, then, though it may be able reliably to classify people as falling one side or another of the cut-off score, this classification will have nothing to do with those people's learning support needs.

Typically, screening tests are designed to err on the side of caution, by having the cut-off score set so that you tend to generate false positives rather than misses. This is because you can always carry out further assessment of those with positive outcomes in order to confirm whether they really do have a support need and, if so, what sort of need it is. However, once you have classified someone as a 'negative' test result, then you have said they are all right. Any needs they might have will have been missed.

So, the purpose of a screening test is to 'catch' as many as possible of those who are likely to need support, even if, at the same time, you also catch a few who do not.

Without going any further, screening tests can provide useful management information. For example, in FE, colleges can find out what proportion of people are likely to need learning support and hence identify the level of resource necessary to cope with that. In so doing, it is not necessary to identify the individual problems each person might actually have: that can be done later. What is important, is that the necessary level of support resource is provided.

Why do we need to use diagnostic testing?

Defining individual problems and tailoring support to individuals requires diagnostic assessment. Those who are classified as 'positive' on a literacy screening test may have a variety of greater or lesser types of problem or support needs. Diagnostic tests can be used to check that they really do have support needs (i.e. that they are not just false positives) and the nature of those needs.

Not only is diagnostic testing expensive in itself, but it has implications for the institution. If you are going to spend the time and money needed for diagnosis, you need the resources to provide the specialised support to deal with the problems you diagnose. If that support is not available, diagnosing the fact that someone has a problem is of little benefit. It is like going through a series of medical tests, being told what is wrong with you, and then being told that there is no means of treating you!

Using screening tests and diagnostic assessments

Screening is used by most colleges to give an indication of levels of basic skills for large populations of students. Cut-off scores are designed to support the process. The cut-off score is the significant factor which can vary the outcome of a screening test and moving a cut-off score up or down obviously affects the proportion of a population that will be identified through the use of a screening test.

It is important therefore that guidance given with screening tests is followed and that the outcomes of testing are interpreted in line with the associated guidance. It is important that tests are administered in the form in which they are presented and are marked and interpreted in line with the guidance provided. The outcomes of these tests provide useful information which can be used:

- at a strategic level to inform decisions about resource allocation and curriculum planning

- as evidence of support needs in applications for additional funding units
- to signal the need for further assessment to identify more specifically the capabilities and needs of an individual in relation to the learning demands which will be made of them.

Screening is only effective when the proportion of people you are trying to detect is significant, and the test you are using is valid and robust. If either no-one or everyone had the quality you were screening for, the test would be a waste of time. It only makes sense to use a screening test when you need to separate out those who do have some quality from those who do not. A particular test may work very effectively as a screening device for an intake population which has around a 50% occurrence of people with basic skill problems, but work less well for one where the occurrence is either much higher or much lower. If the percentage of people in the intake population is either very high or very low, then you need a more powerful test than if the percentage is around 50%.

As a result of this, a particular test could be highly effective as a screening device for one college, which had an intake with a high proportion of people with basic skills problems, but ineffective for another college where the intake was from a very different population.

There are many instances where college staff have devised tests for themselves. These tests are usually programme specific and are sometimes used for screening purposes, sometimes for diagnosing individual needs. Colleges have come to describe these home-grown assessment tools as 'diagnostic'. However, they are often used in the absence of any code of practice and without clear guidelines regarding the training needed for those who use and interpret them. It is important that appropriate guidelines are followed so that the quality and effectiveness of the use of these tools can be assured.

FEDA has collected examples of home-grown diagnostic assessment tools developed by 10 cutting edge colleges, including a few who sell their assessment materials to other colleges and were identified by recent national research conducted by NCVQ as having a significant market share. All the materials and procedures associated with their use were examined and evaluated by the second author (a Chartered Occupational Psychologist with specialist skills in test development and test use). They were found to be of varying quality. Particular problems included:

- setting cut-off scores for classifying people into different levels of attainment without any empirical justification for the location of the cut-off points
- inappropriate diagnostic interpretation of responses to individual items or small subsets of items on tests designed as screening tests
- home-produced tests lacking any evidence of reliability or validity or other supporting technical documentation
- application of inappropriate 'summative' educational assessment approaches to diagnostic assessment.

To help colleges which choose to invest development time in producing their own programme-specific assessment tools, FEDA has distilled from the evaluation of existing materials, good practice criteria and guidance to support colleges in producing more rigorous and robust assessment instruments.

4 Tools and techniques

Quality criteria

For any particular assessment problem, a range of assessment methods may be available. How do you decide which to use? The quality criteria described in *Managing assessment* (FEDA, 1995) can be used as the basis for defining six important factors:

- scope
- reliability/accuracy
- validity/relevance
- fairness
- acceptability
- practicality.

While applicable to all types of assessment, these are of particular importance for forward-looking assessment which is used to make decisions about a person's future. Issues of equity and fairness become paramount in these situations.

Scope: what range of attributes or skills does the test cover, and what range of people is it suitable for?

Assessment methods vary in both their breadth and their specificity. A basic skills screening test may be regarded as 'broad' and 'general'. A battery of tests for the diagnosis of dyslexia would be broad but specific, while a test of punctuation would be narrow and specific.

Using the zoom-lens analogy mentioned earlier, the coverage of the test or test battery refers to how much of the total picture it covers. If it is a specific diagnostic battery, it may provide a set of detailed pictures which cover the whole scene or only some parts of it.

In considering the scope of a test, we also need to ask what populations or groups of people it is intended for. Anyone and everyone, or for those without formal educational qualifications? Is it all right to use it with people for whom English is a foreign language?

Reliability or accuracy: how precise is it?

What reliance can you place on the score somebody obtains? If they did the test again tomorrow, or next week, would you expect them to get the same scores? For objective tests, this 'precision of measurement' is referred to as reliability.

Reliability is assessed in two main ways: by measuring the consistency of the score and its stability.

We can say that a test provides a consistent measure of an attribute if the accuracy of the responses given to each question in the test is related to the 'amount' of the attribute a person has. A test will be inconsistent if there are some questions in the test which require skills or attributes other than those which the test is designed to measure.

A test is 'stable' if a person tends to get about the same score each time they take the test.

Both consistency and stability are usually expressed as correlation coefficients. These range from zero (meaning you cannot place any reliance on what the test score tells you) to one (which means that the score a person obtains is a perfectly accurate measure of the amount they possess of the attribute). In general, we expect tests to have reliabilities of at least 0.70, and up to around 0.90. Values in the low 0.80s are considered good.

Reliability is important because it tells us how much confidence we can place in a score. Any measure of a person's performance has a margin of error associated with it. If you measure someone on a number of occasions, for example, you will get a variety of scores. The extent to which these scores vary from each other (assuming the person's skill has remained constant) is a function of the level of error in the measurement process. The reliability of a test enables us to specify that level of error and to define the width of the 'band' or region within which we can expect a person's true score to lie (as opposed to the actual score they obtain on one occasion). The narrower or tighter this band, the more reliable the test.

Speed versus power tests

Tests of ability and achievement fall into two main types: speed and power tests.

Power tests are those for which enough time is given for most people to attempt all the items, and where the main factor determining whether someone gets an item right is the difficulty of the item.

Speed tests, on the other hand, consist of a number of relatively easy items which have to be attempted under a stringent time constraint. As a result, the main factor determining a person's score is how many items they attempt in the time available, rather than how difficult the items are.

It is very important not to mix up these two types of test. Setting a tight time limit on a power test will render it invalid, as will relaxing the time limit on a speed test. In practice, many tests fall somewhere between these two extremes. Most published tests have been designed for use with time limits, and all the information provided about them is based on people completing them within these limits.

The more a performance is constrained by time, the more difficult it is to get good measures of reliability by using internal consistency. In general, if tests are speed tests, or are power tests operating with tight time limits, re-test correlations will provide better estimates of reliability than measures of internal consistency.

For basic and key skills assessment, power measures are generally more appropriate than speed ones. This does not mean that these tests do not have time limits but that the time limits are intended to be generous enough not to penalise people who are capable but slow.

Reliability and test length

Assuming you could generate unending numbers of questions for a test, all of which related to the same attribute, then the reliability of your test would be a simple function of how many questions you included in it. The longer the test, the more reliable; the shorter the test, the less reliable. This is why you should not use a screening test for diagnosis. As you break down the items in the scale into subscales, so the reliability decreases. An instrument with a respectably reliable overall scale based on 20 questions would be quite useless as a diagnostic tool if you start to look at performance based on subsets of four or five items each.

Reliability is the key to good assessment. You cannot have a 'good' test which is not reliable – though you can have reliable tests which are no good for a variety of other reasons.

Validity or relevance: does the test measure what it claims to measure?

Validity concerns questions such as:

- Is there evidence to show that the scores relate to the qualities which the test was designed to measure?
- Do the scores enable you to make relevant judgements about the person, their current and future performance, their support needs, and so on?

If the measures are unreliable, the answer will be 'No' to both questions. If the measures are reliable, the answers may be 'Yes', depending on the evidence for the validity of the test.

It is useful to distinguish between construct validity and criterion-related validity. For a literacy test, for example, construct validity evidence would be that which supports its general claim to measure literacy. However, the relationship between scores on the scale and, say, future measures of learning support need, is evidence of criterion-related validity.

Construct validity is a must: we can argue that we have a measure of literacy if, for example, we are able to show that it relates to some other measures of literacy. On the other hand criterion-related validity is 'optional'. We may have a very well-designed general basic skills test with good construct validity, but find that scores on its scales do not actually relate to any of the learning support or other criterion measures which we have. Such an outcome does not mean that the instrument is no good – only that it would have no use in terms of the particular external criteria we were considering.

In judging an instrument's worth in terms of criterion-related validation studies, therefore, one has to consider very carefully the relevance and accuracy of the criteria used. The questions we need to ask are:

- Does this instrument measure the characteristics it claims to measure?
- To what learning criteria might we expect such characteristics to be related?
- What empirical evidence is there to show that these relationships actually exist?

Evidence of criterion-related validity (or criterion relevance) comes in a number of forms.

- The test could have been given to a large number of people whose learning support needs were subsequently monitored and recorded. Some time later this information would have been related

back to their test scores. The relationship between their performance on the test and their subsequent support needs would be a measure of the predictive validity of the test. This is the strongest form of evidence for relevance.

- A similar idea involves obtaining test scores from a group of people who may be following a training programme, and who vary in their current support needs. The relationship between their test scores and the measures of support is called concurrent validity (as both are being measured at the same time). A problem with this approach is that if the learning support has been effective, then taking both test and learning support measures at the same time will give misleading results. The more effective the learning support, the less predictive the test would appear to be.

Other forms of validity are matters of judgement rather than objective 'fact'.

- Expert judgement as to the relevance of the content of the test items for various different purposes is known as content validity. As experts can be wrong, this should never be taken as convincing evidence of relevance on its own.
- Face validity is simply what the person taking the test thinks the test is assessing. Face validity is important in establishing a good rapport with the candidate and ensuring co-operation in the assessment process. However, good face validity is no guarantee that a test is actually either relevant or useful.
- Finally, there is faith validity. This concerns the test user's beliefs in the value of the test – generally in the absence of any evidence to support it. This is one of the greatest problems to counter. It is belief or faith, rather than sound evidence, for example, which is the basis for the continued acceptance of techniques like astrology and graphology in some areas of assessment. There is a very natural tendency to over-interpret assessment data (from all sources); to see patterns where none exist. The technical information provided with good tests is intended to 'restrain' users from doing this.

Fairness, or freedom from systematic bias

Are the results for different groups of people likely to differ systematically for reasons which have nothing to do with the relevance of the test? It is important when looking at any instrument to check for information relating to bias and possible sources of unfairness. Information should be provided about differences relating to gender and relevant minority groups. Ideally, information will also be presented to show that the individual items have all been examined for bias – by careful examination of their content, and by statistical item-bias analysis.

However, it is also important to remember that bias is not the same thing as unfairness. Statistically, bias simply means that there is a systematic tendency for members of one group to respond differently from those of another. Whether or not that is unfair is quite a different question.

Suppose we wanted to use a test to assess the level of learning support provision needed for a particular training course. We use a test on which a score of less than 65% correct indicates a need for specific support. Suppose also that we find, on average, that women obtain lower scores than men on this test. This would suggest that women were more in need of learning support than men. However, there are two possible reasons for this outcome.

This is an accurate reflection of a gender-related difference (in which case the test is a 'fair' reflection of the real world – even if the real world is not 'fair').

There is a systematic bias in the *test* which results in women and men with equivalent learning support needs obtaining different scores. If this were the case, the test would over-estimate the number of women and under-estimate the number of men needing learning support.

In general, if it can be shown that the differences between any groups of people (either men and women, different ethnic groups, or old and young people) reflect real differences in their performance, then any test bias is 'fair'. If, on the other hand, the relationship between test scores and reality differs from one group to another, then the test is 'unfair' if these differences are not compensated for in some way.

It is very important when planning to use a test to check for evidence of bias (are the average raw scores different for males and females; do they vary with age, ethnic background etc.?). In particular, great care should be taken when assessing people with disabilities to ensure that their disabilities will not unfairly impact on their test performance.

Numeracy and literacy

Before you use tests on anybody – especially tests of basic skills – you should check that the person has the numeracy and literacy skills needed to understand what is required to carry out the test. Those

for whom English is a second language may have language problems which interfere with their potential to respond, for example, to numeracy measures.

The same is true of numeracy difficulties: if a test is designed to measure some attribute other than numeracy, but requires the test taker to be numerate in order to do the test, then it would have an unfair impact on those with numeracy problems.

Setting a level playing field

It is important to ensure that when a group of people sit down to take a test, they have all been provided with the same information about what they are about to do, and have been given the time and opportunity to become familiar with the testing process and the sort of materials they will be dealing with.

To ensure that there has been no inequality of opportunity in terms of fore-knowledge and preparation, most test publishers now produce practice tests or practice leaflets. These are intended to be given to test takers well in advance of their test session. It is also good practice to provide them with an opportunity to ask questions about the procedure before the test session starts.

Information provided before the test session should also make clear how you intend to use the information collected, who will have access to it, and for how long it will be retained. Following a code of good practice (see Appendix 1) will help to ensure that procedures for testing are fair and consistent.

Acceptability: can you expect people to co-operate in the assessment procedure?

In general people find tests acceptable when:

- the reasons for taking the test have been carefully explained to them
- they have been given adequate prior information about the nature of the assessment and the opportunity to ask questions
- the administration is properly carried out
- they are provided with feedback about their results and their implications.

Both faith and face validity, discussed earlier, affect acceptability. Acceptability is important because it affects the degree of co-operation you can expect from the test taker and the rapport you can establish with him or her in the assessment process.

Practicality: what does it cost, how long does it take, what equipment is needed?

The quality of the information which an assessment procedure provides must be weighed against the cost of obtaining that information.

Objective tests are highly cost-effective as they provide a lot of accurate, relevant information in a short space of time.

The results of objective tests provide information which it is very difficult to obtain using other methods.

However, there are costs. These fall into three areas:

- the costs associated with becoming competent as a test user
- the cost of buying or developing test materials
- the recurrent costs of using tests (time and materials).

You need to be properly trained and qualified if you are to use the tests appropriately and if you are to provide fair and balanced interpretations of their results. Improper use of tests and test results – or other less objective methods – is not only potentially damaging to the individuals concerned, it can also be a waste of resource from the organisation's point of view and, increasingly these days, carries a risk of litigation. It is far less expensive to make sure you are competent in the first place, or that you seek advice from competent experts.

Any college which is seriously involved in diagnostic assessment should have at least one person on the staff who understands the principles and basic technicalities of objective measurement. These are spelled out in the 'Level A' test user standards specified by the British Psychological Society. Anyone intending to develop their own screening or diagnostic instruments, will need a higher level of expertise than this, or should seek the support of an outside test development specialist.

What resources are required?

Objective test materials can seem quite expensive. As a result, it can seem like a good economy to create one's own. However, their price is a reflection of the development costs associated with producing instruments which meet the criteria discussed above. Furthermore, national and international publishers

can make economies of scale which are not open to individual colleges. Making up your own tests is likely to be a false economy. Copying other people's without permission is illegal.

The indirect costs associated with maintaining an assessment resource also need to be considered. These include provision for storage of materials, time, and the cost of developing and implementing an organisational testing policy and carrying out quality control procedures. If you are only likely to carry out the occasional objective assessment then it is probably more cost effective to use qualified chartered psychologists as consultants. Where you have a need for regular use, you need to set the costs (in terms of initial training, materials purchase and other set-up costs) against the benefits which will accrue.

While there is little doubt that good objective tests, properly used and interpreted, are one of the best sources of information about people's attainments and potential, there are costs associated with them:

- Becoming a qualified test user involves training which can be expensive and takes up valuable time.
- Establishing a suitable library of test materials requires a financial investment (although the actual per candidate cost of testing tends to be quite low).
- Developing one's own tests, as described above, is a high-cost option. It also requires areas and levels of expertise which are likely to lie outside those possessed by staff in FE colleges.
- Test administration and interpretation take up time which you may need for other activities.

Record-keeping, monitoring and follow-up

To maintain quality control over any procedure, you need to record what happens and follow through decisions which have been made to see how effective they were. It is thus important to keep records but also important to ensure that the information they contain remains confidential.

You should keep a record of which tests you have used, when you used them and why.

You will also need, with the test taker's permission, to keep a record of test results while they remain your responsibility.

As a rule of thumb, it is a good idea to keep in mind that any information you might obtain using an objective test 'belongs' to the test taker. Whatever you do with it should, therefore, only be done with their knowledge and permission.

Actual test scores should only be passed on to people who are qualified to interpret them. This includes your students. Otherwise, test interpretation in plain straightforward English is all that should be given to other people. If possible, test scores and general information about the person's age, gender, educational background etc. should be archived for possible future use in test refinement and development, the production of new norms and validation.

However, if you do store test information for such long-term purposes then, for the protection of the test takers, you should be very careful to ensure that you do not keep any information which might enable individual people to be identified.

What are the costs and benefits of diagnostic testing?

In general the resources needed to do testing well are:

- time
- money
- expertise
- access to relevant populations.

Of these, only the last is readily available within FE colleges. If the development of testing and assessment procedures results in an overall net efficiency gain, then it is worth pursuing. The costs are high, up-front, and very apparent. The benefits can be more difficult to measure and more difficult to attribute directly to the use of diagnostic testing.

In considering cost-benefit trade-offs, we need to look at the possible outcomes associated with using tests in this way (both positive and negative), and the consequences of not using tests.

Bad diagnostic tests result in mis-diagnosis. This is likely to be a more expensive mistake – in both financial and human terms – than no diagnosis at all. Good diagnosis, on the other hand, can result in the efficient targeting of scarce resources (staff time). It will help ensure that students are given the support they need and are not de-motivated by either being 'helped' inappropriately or unnecessarily, or failing to cope with the demands of their course.

Costs include the design and development costs (start-up costs) and the recurrent costs. Design and development costs can be high. As the previous

section illustrates, developing these instruments is not a simple process. Costs can be reduced by buying in existing materials. However, if you do this, you need to have good technical evidence to support their quality.

It is necessary to see these costs as part of the overall costs associated with delivering good learning support. As pointed out before, diagnostic testing is pointless unless there is appropriate, competent 'treatment' available.

KEY POINTS

- Quality control in the use of objective assessment depends on the combination of a robust, relevant instrument and a competent user. The user has to be trained to understand the instrument and to use it appropriately within the limits of its technical characteristics. The instrument needs to be fit for its intended purpose.
- In making this judgement, six main areas of quality control criteria have been highlighted: scope, accuracy, relevance, fairness, acceptability and practicality. To judge the overall cost-effectiveness of using any assessment method, one should evaluate it against all six of these criteria.
- Accurate diagnosis is not the same as effective 'treatment' – but it does provide the information needed to target support resources more efficiently.
- Overall efficiency needs accurate diagnosis, combined with appropriate placement and good learning support.
- Poor diagnosis can be costly to develop and lead to a misdirection and waste of support effort.
- Poor diagnosis can end up costing you more than no diagnosis.
- Consider the costs and benefits of alternative strategies to diagnostic testing, for example, screening with flexible learning support. We should always ask the question:

  What value would diagnostic testing add to that obtainable through screening tests on their own?

BEST COPY AVAILABLE

5 Advice to colleges: dos and don'ts

Assessment policy

Do develop and adhere to an organisational policy on assessment.

Where initial assessment involves diagnostic testing, the policy should embody the principles represented in the attached draft Code of Good Practice in Testing (see Appendix 1) and cover in detail:

- test supply and control of materials
- what tests are used
- how they are used
- who is responsible
- what training and evidence of competence are required of test users
- what limits are set on their use of testing
- how results are to be given to students and other support staff
- how data are to be stored
- what provisions are made for monitoring the quality of testing
- what procedures have been put in place for following up the effectiveness of testing (to assess cost-benefits).

Making a contract

The students you test need to understand why you want to test them, what the process involves and what they will get from it. It is a good idea to establish a form of 'contract' with them which makes clear, from the start, how your policy is implemented.

What are the options?

If they agree to take tests, what assurances do they have that the tests used will be good ones, and that the people doing the testing will be competent?

What will happen to their results and who will have access to them?

What support will be available for them if the tests suggest they need it?

What are their rights and responsibilities in the process?

To test or not to test

Do use screening. Screening is relatively inexpensive and good tools are available. The incremental benefits of diagnostic testing may be relatively small compared with the costs.

Do not use tests as the sole basis for diagnosis and guidance.

Do not test people for the sake of it.

Which tests to use

Do use tests of known quality.

Do not be fooled by appearances: testing materials may be well presented without being rigorous and robust testing instruments.

Do not buy tests on the basis of personal recommendations or testimonials. It is the technical evidence which you need to consider.

Do-it-yourself?

Do not just 'buy in' from another college in order to save time and money. Another college may have spent a lot of time developing some tools, but you do not want to be paying for their wasted time if the tools are no good.

Do think long and hard about the costs associated with developing your own instruments before you get involved in doing so.

Do seek expert advice. Test design, development and validation is a highly technical complex process. Even the experts find it difficult to develop good tests.

6 Conclusion

Those working in FE colleges face a dilemma. It would appear that there is no easy access to commercially produced, good, diagnostic materials at present. Yet FE teachers perceive a need for diagnostic tools to help them tailor support and to advise students better on their options. Well-designed diagnostic tools which will give people the information they want, however, are expensive to develop and are unlikely, in practice, to be a cost-effective option given the time and money involved in their development against the increase in quality of provision they may afford.

Many people have produced assessment materials. While some of these may provide reasonable assessments of attainment in relevant areas, they are not developed as 'forward-looking' measures. The danger lies in the results of such assessments being used as if they were objective diagnostic tests.

FEDA is doing new work, funded by the Qualifications and Curriculum Authority, to develop tools for the initial assessment of key skills. Dave Bartram and the University of Hull are involved, and good practice criteria and guidance published in this report will, of course, inform the work.

If this guidance has done nothing more than make clear the complexity and difficulty of the task involved in developing initial assessment procedures which meet the quality criteria we have set out, then it will have succeeded at least in part. Hopefully, it will also have provided some help to those who wish to ensure that their assessment procedures meet best practice, and are effective. Increasing the effectiveness of such procedures will ultimately benefit all concerned: individual students and staff, the college, and the wider community.

Appendices

Appendix 1: Draft code of good practice testing

For incorporation into an organisational policy on assessment

People who are responsible for the use of diagnostic tests must:

Responsibility for competence

- take steps to ensure that they are able to meet all the standards of competence defined by the British Psychological Society (BPS) for the relevant Certificate(s) of Competence in Occupational Testing, and endeavour, where possible, to develop and enhance their competence as test users
- monitor the limits of their competence in objective testing and neither offer services which lie outside their competence nor encourage or cause others to do so

Procedures and techniques

- use tests only in conjunction with other assessment methods and only when their use can be supported by the available technical information
- administer, score and interpret tests in accordance with the instructions provided by the test distributor and to the standards defined by the British Psychological Society
- store test materials securely and ensure that no unqualified person has access to them
- keep test results securely, in a form suitable for developing norms, validation, and monitoring for bias

Welfare of test takers

- obtain the informed consent of potential test takers, making sure that they understand why the tests will be used, what will be done with their results and who will be given access to them
- ensure that all test takers are well informed and well prepared for the test session, and that all have had access to practice or familiarisation materials where appropriate
- give due consideration to factors such as gender, ethnicity, age and educational background in using and interpreting the results of tests
- provide the test taker and other authorised persons with feedback about the results in a form which makes clear the implications of the results, is clear and in a style appropriate to their level of understanding
- ensure test results are stored securely, are not accessible to unauthorised or unqualified persons and are not used for any purposes other than those agreed with the test taker.

Appendix 2: Defining objective tests

Objective tests comprise a series of standardised tasks (typically, questions to answer, statements to judge or comment on, problems to solve).

Objective tests differ from 'home produced' questionnaires, checklists and observations in that tests are designed in such a way that everyone is given the same task and a standard set of instructions for doing it.

People administering the test are given detailed instructions for preparing candidates, administering the task and for scoring and interpreting it.

Because tests have these properties they allow the trained test user to make objective statistically-based judgements and predictions about a range of issues. For example:

- a person's capacity or potential to act or behave in certain ways
- the likelihood that they will be able to cope with the demands of a training course
- their potential for success in certain types of job.

Objective tests fall into two broad categories: measures of maximum performance and measures of typical performance.

Measures of maximum performance measure how well a person can perform. They have right and wrong answers.

Measures of typical performance measure people's preferences, styles and modes of behaviour. They measure their interests and personality; their values and what motivates them; the attitudes and beliefs they hold.

The present guide is only concerned with the first of these.

Tests of maximum performance include general ability tests. These provide a good indication of a person's potential to succeed in a wide range of different activities. Such measures are relatively unaffected by the person's previous experience and learning.

Measures of attainment or mastery, on the other hand, specifically assess what people have learned and the skills they have acquired (e.g. shorthand and typing tests; knowledge of motor mechanics and so on). Where these focus on very specific aspects of skill (e.g. punctuation, forming plurals, and so on) they are often called 'diagnostic tests'.

Between these two extremes there are a number of other types of tests: specific ability tests, aptitude tests, and work-sample tests.

Objective tests provide a means of comparing an individual against known benchmarks (typically, the average performance of some defined population, or explicit mastery criteria).

The actual number of correct answers a person gets on a test is known as their raw scale score. There are, however, a number of other types of score typically used in testing. These various different measures are referred to as standardised scale scores.

For some tests, all the correct items are counted together to produce just one scale score. For other tests, the scoring procedure may divide the items into two or more groups, with each group of items being used to produce its own scale score.

So, what is a scale? To give an example, we might have a screening test which contained 30 items designed to measure literacy and 30 items to measure numeracy.

If the two sets of items are presented as two separate tests we would call it a test battery (containing two tests: one for literacy and one for numeracy) and we would have two scale scores.

Alternatively, we might use the test to provide a general measure of basic skills. In that case all 60 items would be used to produce a single scale score.

A further possibility would be for the 30 items in each test to be broken down into sub-scales, each measuring some distinct aspect of literacy (spelling, use of punctuation, etc.) or numeracy (fraction to decimal conversion, multiplication, etc.).

The extent to which we can break down a scale into meaningful subscales depends on how the instrument was constructed, and what degree of accuracy we need in our measurement. In general, the more you break scales down into components, the less accurate the component measures become.

There is a very real danger of people taking tests designed for screening purposes and then over-interpreting them. For example, a general screening test for numeracy would probably contain items covering addition, subtraction, multiplication, division, conversion between decimals and fractions, use of percentages, and so on. This does not mean that you can look at a person's performance on each type of item and diagnose their relative strengths and weaknesses. You can only do that if the test was designed so that each item type produces a scale score (of known reliability and validity).

Comparing scores with other people's scores: using norm-referenced scores

A norm-referenced score defines where a person's score lies in relation to the scores obtained by certain other people. This group of other people is known as the norm group. The reason for using norm-referenced scores is to see whether the person tested is below average, average or above average with respect to the performance of the norm group. Such scores are relative measures as they depend on who the 'other people' are. For example, a numeracy score which may be average for one group, may appear 'low' when compared against a university graduate norm group, and 'high' when compared against a sample of people drawn from a population of school leavers from a depressed inner-city area.

Typically norm-referenced scores are expressed either as percentiles or percentile-based grades or on one of a number of standard score scales (e.g. sten scores and T-scores), for example:

The Basic Skills Test (NFER-NELSON) uses normalised T-scores and percentile scores (and gives 68% true score confidence bands).

The Foundation Skills Assessment (Psychological Corporation) provides percentile points, Stens, Stanines and ratio scale scores which allow comparison between the different levels (A, B and C) of the test.

Comparing scores against a standard: using criterion-referenced scores

To make judgements about a person's ability to cope with the demands of a job or training course, test scores have to be related to performance on the 'criterion' task – e.g. the training course. This is typically done in one of two ways.

In one approach, 'experts' make judgements, based on an examination of the content of the test items and of an analysis of the demands which will be made on people by the course, about what minimum test scores would be required for a person to be able to cope with particular aspects of a training programme. This method is variously referred to as domain-referencing, content-referencing (and also, confusingly, criterion-referencing). This is the most common approach, widely used in educational assessment, and likely to be the method used by most FE Colleges. The Basic Skills Assessment (The Basic Skills Agency) adopts this approach producing raw scores which are criterion-referenced in relation to three stages: Foundation, Stage 1 and Stage 2.

The second method involves making statistical predictions of future performance from a person's scores (called predictive validation). This uses actual data about the relationship between test scores and educational attainment or training course performance. Establishing this sort of relationship is especially important when the results of tests are used in the process of selecting people for jobs or training courses – as opposed to post-selection evaluation of their support needs.

Objective tests have quantified levels of accuracy of measurement and a body of evidence supporting their claim to measure what they say they measure.

One of the properties which distinguishes objective assessment from other forms of assessment is the quality and quantity of the data available about the instrument. The interpretation of any scale score is aided by:

- information about how scores are distributed in the general population and in other more specific groups of people
- information about variations in patterns of score distribution related to demographic variables – such as age, gender, ethnic group and so on
- information about variations in patterns of score distribution related to occupational or educational criteria.

Information on the effects of demographic variables is very important in judging the suitability of an instrument and in aiding interpretation of it. One should always look for data on these variables, for example, whether scores on each scale differ between males and females or not, whether they are subject to ethnic group bias effects, whether the instrument has been used on populations similar to those you would be assessing.

Appendix 3: How are diagnostic tests designed and developed?

This appendix focuses on the development of diagnostic tests. It assumes that there is already a screening testing process in place identifying people with probable learning support needs. Diagnostic tests are then used to specify in more detail the nature of those needs and the extent to which they will require support.

In principle, the process of developing diagnostic tests is straightforward. The difficulty and complexity lie in the details of doing it properly and emerging with a robust and useful instrument. Specialist technical assistance will be needed to do this properly. Most of the technical difficulties arise towards the end of the process (and relate to data analysis and technical documentation issues). However, you are strongly advised to seek the advice of a chartered psychologist with specialist skills in test development from the start. They will be able to guide you through the initial stages so that you stand the best chance of producing an analysable and useful set of data at the end.

Profile the demands which the programme will make on the learner

Break down the course programme content in terms of the demands the course materials and content will make on the learner. As well as helping to define levels below which support will be needed, this also provides an opportunity for checking whether these levels of demand are appropriate for the course. For example, could the course objectives still be met if difficult materials were revised to make them less demanding of basic skills? Costly learning support provision could sometimes be reduced by changing the demands made on learners by the course.

Define the essential learning skills required to cope with these demands

Having profiled the demands, identify the skills required to cope with these: for example literacy, numeracy, communications, IT, etc.

Identify the minimum necessary levels of skill needed in each area

In each area, ask what the minimum necessary level of skill would be for someone to be able to cope with the course content. This should be the level below which the person's lack of skill will start to get in the way of, or interfere with, their ability to keep up with the course and maintain steady progress.

Devise or choose tasks which will assess the skills concerned

Having identified the relevant areas of skill, choose tasks which will assess those skills. This may be obvious in some cases (e.g. coping with fraction to decimal conversions), less so in others (levels of literacy needed to understand the range of reading materials which accompany the course work).

Set the difficulty of the tasks for optimum discrimination around the minimum necessary level

In designing the tasks for the test, aim at the people on the borderline of the level of competence needed to cope with the course. They should be able to get about 50% of the questions right (assuming open-ended answers, or multiple choice with corrections for guessing). If the test is too easy or too hard, it will not identify the deficits you are looking for. Getting this right is very difficult. A common mistake is to make the test representative of the course materials. This is wrong because much of the course may not be problematic. The test should be selective – focusing on areas where there is likely to be difficulty for those identified as 'at risk' by the general screening process.

Ensure the content of the items has sufficient scope to cover the range required

This is a similar point to the one made above. The items, or questions, in the test need to cover all the relevant areas. This is vital for diagnostic instruments. Never assume that there is a problem in one area simply because you have diagnosed a problem in some other area. People are likely to differ considerably in the extent to which their learning support needs are general or specific.

Ensure there are sufficient items to give a reliable measure

The main shortcoming of the 'home-grown' variety of diagnostic tests is that they do not contain sufficient items for a reliable diagnosis to be made. It is not uncommon for tests produced in colleges to contain only one or two items relating to a diagnostic category. This is quite insufficient to diagnose problems at an individual level. As a rough and ready rule of thumb, the minimum number of items per diagnostic category should be between 6 and 12. The number really depends on how narrow or broad the category is. For very narrow, highly specific ones, you may be able to produce reliable scales with only six items.

A related issue is that when devising your own materials, you need to make provision for the fact that some of the items you produce will not work as you

expected them to. Most test developers reckon on trialing as many as twice the number of items they expect to end up with in the final test. So, if you are devising a test to diagnose difficulties in six different but fairly specific areas, you would probably need to generate over 100 items in the first instance.

Some forms of test do not use 'items' in the conventional sense of the word (that is discrete questions). You may instead use passages of text about which a number of questions are asked; you may ask people to write free-text which is then content-analysed; and so on. While it is important that the form and content of the test materials are relevant to what is being assessed, the issue of reliability and robustness remains. For even the most open-ended form of assessment, you need a well-defined scoring protocol which determines what 'items' of data are obtained. These item scores should be treated, for statistical purposes, just as the scores you would obtain with closed, multiple-choice test questions.

A final factor to consider in designing test items and ensuring there are sufficient, is inter-dependence. If you present people with a short passage of text, and then ask 10 questions about it, there is likely to be interdependence between the accuracy of the answers because they all relate to a common 'stem'. Technically, this means that a measure of the reliability of this test will be artificially inflated. An extreme example would be asking the same question 10 times. You would have a highly reliable, but not very useful measure. So, one needs to be careful about how far responses may be inter-dependent: for example, where getting Question 1 wrong implies you are likely to get Questions 2-10 wrong as well because they all relate to the same material.

Pilot test the items using people with known learning support needs or known levels of literacy or numeracy

The best way of getting data on whether the test is right or not, is to try it out. The better defined the sample of people it is tried out on, the better the quality of the information obtained. At this stage you are not looking for large numbers of people – that comes later. You do, however, want to be able to check that the types and levels of skill you are intending to measure can be identified in people already known to have those levels and types of skill

.Ask people with relevant expertise to check the item domain and difficulty. This can be done using sorting and rating tasks

In addition, you can use expert judgement to help check out the item design. A common approach to this is to get a small group of experts (five or six will do), and ask them to carry out a sorting task.

The task is to sort the test questions into piles, each pile relating to one of the characteristics you are attempting to measure. Suppose you were trying to diagnose problems in three aspects of numeracy, and had designed items to be appropriate to 'Foundation' level skill. You provide your experts with a matrix in which to sort the items: the rows would be difficulty levels: 'Foundation', 'Level 1', and 'Level 2', say. The columns would be heading with the various aspects of numeracy into which the items could be classified – together with a 'Don't know' column. Each item in the test is written on a separate card, and the experts are given a pack of cards and asked to place them in the appropriate cells in the matrix. Your experts can be course team members, or others who would be in a position to make the necessary judgements.

When this has been done by all your experts, you need to see whether there is clear agreement between them. Items placed in the 'Don't know' column by more than one person, or those placed in more than one cell should be looked at very carefully or dropped.

Trial the test

Having carried out the pilot testing and the expert sorting, you will now have fewer items than you started with. (For example, an initial set of 100 questions may be down to about 70). These now form the basis for getting some real data under proper testing conditions. In order to do any useful objective appraisal of the test, you will need to get around 100 people (preferably more) from the target population – that is, those people with whom the test is to be used.

Do item analysis to see if the items are working as intended

This is where you may need to call on specialist help. However, there are some simple things you can look at. For example, what proportion of people get each item right? Any item which is got right by nearly everybody or by hardly anyone, is not going to be of any use, as it will not enable you to discriminate between people. So, you would normally discard items which have very high or very low scores (typically, more than 90% correct or less than 10% correct). The exact criteria depend on a range of

matters: for example, what sort of items they are, what the guessing rates are, what the constitution of the trial population was. This last point is very important. If you trial the test with a sample of people who do not have the relevant learning difficulties, then they will be likely to have very high scores on all the items. On the other hand, if your trial group only contains people with these problems, the items will have very low scores. The normal guidelines, then, only apply when your trial group is a reasonable mixture of people with and without the target problems.

Other scale-construction processes which need to be carried out at this stage are reliability analysis, examination of item discriminations and, possibly, principal components analysis. These require specialist software and specialist knowledge. It may also be possible, depending on how your trial sample of people is constituted, to carry out a preliminary validity analysis. If you know which people had and which did not have the target difficulties, you can see what proportion of them were correctly identified by the test. Again, this sort of analysis requires specialist help. Where target groups are well defined within the trial sample, discriminant function analysis can be used to develop prediction scores.

Revise the test this stage is to drop items which are redundant or do not work. This is done on the basis of the scale construction analysis work. All the final validation and reliability analysis is done on the revised test – not the original set of items which were used.

Establish criterion points and cut-off scores

It may be possible to set provisional cut-off scores on the new test using the trial sample data – or it may be necessary to get more data, using the final version of the test, to do this. Again, you may need specialist advice on this.

Document the technical information

Once all this work has been completed, you must ensure that the technical information about the test is properly documented. This will normally be in addition to, and separate from, any documentation you might produce for those staff using the test on a day-to-day basis. It is doubly vital to produce good technical documentation if you intend offering the test to other institutions.

Producing technical documentation is a specialist job, and one which staff in FE colleges are unlikely to have the necessary expertise for. It is a task outside the range of competence of the average test user.

Follow use of the test

Once the test has been developed, it is vital that its value is assessed. This means following all the outcomes:

- How accurately does it identify programme specific support needs?
- Is it adding anything to the information obtained from screening or other sources?
- What are the false positive and false negative rates?
- Would it be worth investing further time and effort in making improvements?

The answers to these questions will all add to the value of the instruments and should be part of the technical documentation. If you are planning to sell your materials to others it is even more important to make sure you have evidence to support whatever claims you make about them.

If you are going to distribute your test materials (either free of charge or for gain) you need to make clear to potential users how the tests should be used and what they can and cannot do with the materials (in terms of making changes, adapting them for their use, photocopying, passing them on to others, etc.).

BEST COPY AVAILABLE

References

Further Education Development Agency (1995) **Managing Assessment.** FEDA, Blagdon.

Further Education Funding Council (1996) **Inclusive Learning: report of the Learning Difficulties and/or Disabilities Committee. FEFC** The 'Tomlinson' Report

Further Education Funding Council (1997a) **How to apply for Funding 1996-7. FEFC.**

Further Education Funding Council (1997b) **Measuring achievement: further education college performance indicators 1994-9. FEFC**

Green, M (1998) **Different approaches to learning support. FEDA.**

FEDA publication series

Developing FE: Volume 1

1 Student tracking
2 Caseloading
3 Assessing the impact: provision for learners with learning difficulties and disabilities
4 Adults and GNVQs
5 On course for next steps: careers education and guidance for students in FE
6 Marketing planning
7 Managing change in FE
8 The effective college library
9 Appraisal in FE – where are we now?
10 Clarity is power: learner outcomes, learner autonomy and transferable skills

FEDA Reports (formerly Developing FE): Volume 2

1 Investing partners: further education, economic development and regional policy
2 Women at the top in further education
3 Moving into FE: the voice of the learner
4 Additional support, retention and guidance in urban colleges
5 Qualifications for the future: a study of tripartite and other divisions in post-16 education and training

FE Matters: Volume 1

1 Environmental education throughout FE 1: policy and strategy
2 Environmental education throughout FE 2: a model and unit of environmental learning outcomes
3 Colleges working with industry
4 Towards self-assessing colleges
5 Evidence for action: papers prepared for FEFC's Learning and Technology Committee
6 Student retention: case studies of strategies that work
7 Getting the credit: OCN accreditation and learners with learning difficulties and disabilities
8 Moving on from Key Stage 4 – the challenge for FE
9 Monitoring student attendance
10 Educational psychologists in further education
11 Assuring coherence in individual learning programmes
12 Adult learners: pathways to progression
13 A real job with prospects: supported employment opportunities for adults with learning difficulties and disabilities
14 Transforming teaching: selecting and evaluating teaching strategies
15 Information and learning technology: a development handbook
16 Delivering Modern Apprenticeships
17 Planning a merger of FE colleges
18 Tackling drugs together: addressing the issues in the FE sector
19 Security is not an option – learning in a safe environment
20 Give us some credit: achieving a comprehensive FE framework

FE Matters: Volume 2

1 Youth work in colleges: building on partnership
2 The shrinking world: international links with FE
3 A sense of achievement: outcomes of adult learning
4 Learning with business: FE staff development to business and industry
5 Beyond responsiveness: promoting good practice in economic development
6 Making learning support work

FEDA Bulletins: Volume 1

1 Quality assurance in colleges
2 The impact of voucher schemes on the FE curriculum
3 Enhancing GCE A-level programmes
4 Developing college policies and strategies for human resource development
5 Maintaining quality during curriculum change
6 Action planning and recording achievement
7 Implementing modular A-levels
8 Comparing content in selected GCE A-levels and Advanced GNVQs
9 Engineering the future: monitoring the pilot GNVQ in engineering
10 Charters in FE: making them work
11 Access to accreditation
12 Back to the future: what are Modern Apprenticeships?
13 Competing for business: colleges and the Competitiveness Fund
14 Information systems: a strategic approach
15 Strategic approaches to processes, cultures and structures

FEDA Bulletins: Volume 2

1 Partners in economic developement
2 Assuring quality in collaborative provision
3 Student stress in an FE college: an empirical study

FE matters

BEST COPY AVAILABLE

ISSN 1361-9977 FE matters Vol 2 No 7 Price £7.50